

Science and religion in harmony

A spiritual leader with an interest in research has encountered opposition to his plans to speak at a scientific meeting. But he is perfectly entitled to do so.

The Dalai Lama is due to speak at the annual Society for Neuroscience meeting in Washington DC on 12 November, and some neuroscientists don't like it (see *Nature* 436, 452; 2005). But the Buddhist leader's talk is part of a lecture series that the society is laudably conducting on the science and society — and it should go ahead as planned.

The invitation of the Dalai Lama to the meeting will be interpreted in some quarters as an insult to his nemesis, China. And, citing the oft-repeated refrain that science and religion should be kept separate, some neuroscientists are calling for the lecture to be cancelled.

The critics accuse the Dalai Lama of trying to use the meeting to sell science that they regard as substandard: research on the relationship between meditation and physiological changes in the brain. Even the researchers directly involved in these studies, many of whom are working with the encouragement and support of the Dalai Lama, say that the work is in its early stages.

But the society did not invite the Dalai Lama to speak as a scientist. He will be in Washington to kick off its lecture series on "Dialogues between Neuroscience and Society", in which non-scientists are expected to address "subjects of interest to neuroscientists". The second such lecture will be given by Frank Gehry, the architect who designed the Guggenheim Museum in Bilbao, Spain.

Since *Nature* first reported on this story three weeks ago, several neuroscientists have written to us criticizing efforts to stop the lecture (see page 912, for example). It seems reasonable to assume that a fair number of the 30,000 delegates expected to attend one of the world's largest scientific meetings will be interested to hear what the Dalai Lama has to say.

The Dalai Lama will not be a complete outsider at the meeting. Through the Colorado-based Mind & Life Institute, he has already interacted with many reputable neuroscientists. According to the society, he was invited, in part, because "he has already had an influence on the design of experiments of great interest to neuroscientists". As even one opponent of the talk admits: "He has views on

controlling negative emotions, which is a legitimate area for neuroscience research in the future." But his lecture does not necessarily constitute an endorsement of his views by the society.

Critics counter that the talk threatens to "entangle the Society for Neuroscience with religious activities". The invitation for the Dalai Lama to speak will give him a chance to sell his religious beliefs in the guise of neuroscience, they claim. Their petition opposing the lecture even draws comparisons between the Dalai Lama, with his belief in reincarnation, and creationists.

But speakers at meetings — non-scientists or scientists — should not be barred on the basis of their religious beliefs. Well-known scientists including Newton have had religious beliefs that many people would disagree with, but these have no bearing on the credibility of their scientific ideas.

Furthermore, in stark contrast with the approach of most religious leaders, the Dalai Lama has tried for many years to encourage empirical research into the claims he makes for the value of meditation. He encourages monks to take part in such experiments. Resulting studies have appeared in respectable scientific journals.

It is true that the invitation could be interpreted as an insult to China. But the manner in which it was issued — by a scientist who was attending a meeting on neuroplasticity at the Dalai Lama's home in India — implies that the neuroscience society harbours no such intent.

It is not unreasonable for the researchers who object to the invitation to protest against it, and to seek to draw attention to the limitations of the Dalai Lama's credentials as a speaker. But now that the point has been made, they should withdraw their threatened boycott of the meeting, and instead raise their issues in the open forum that will follow his talk.

"The Dalai Lama encourages monks to take part in experiments. Resulting studies have appeared in respectable scientific journals."

Ratings games

Researchers have two rare opportunities to influence the ways in which they may be assessed in future.

How to judge the performance of researchers? Whether one is assessing individuals or their institutions, everyone knows that most citation measures, while alluring, are overly simplistic. Unsurprisingly, most researchers prefer an explicit peer assessment of their work. Yet those same researchers know how time-consuming peer assessment can be.

Against that background, two new efforts to tackle the challenge deserve readers' attention and feedback. One, a citations metric, has the virtue of focusing explicitly on a researcher's cumulative citation achievements. The other, the next UK Research Assessment Exercise, is rooted in a deeper, more qualitative assessment, but feeds into a numerical rating of university departments, the results of which hang around the necks of the less successful for years.

Can there be a fair numerical measure of a researcher's achievements? Jorge Hirsch, a physicist at the University of California, San Diego, believes there can. He has thought about the weaknesses of current attempts to use citations — total counts of citations, averaged or peak citations, or counts of papers above certain citation



thresholds — and has come up with the ‘h-index’. This is the highest number of papers that a scientist has written that have each received at least that number of citations; an h-index of 50, for example, means someone has written 50 papers that have each had at least 50 citations. The citations are counted using the tables of citations-to-date provided by Thomson ISI of Philadelphia. Within a discipline, the approach generates a scale of comparison that does seem to reflect an individual’s achievement thus far, and has already attracted favourable comment (see page 900). The top ten physicists on this scale have h values exceeding 70, and the top ten biologists have h values of 120 or more, the difference reflecting the citation characteristics of the two fields.

The author placed his proposal on a preprint server last week (www.arxiv.org/abs/physics/0508025), thereby inviting comment before publication. Given the potential for indicators to be seized upon by administrators, readers should examine the suggestion and provide the author with peer assessment.

Whatever its virtues, any citation analysis raises as many questions as it answers and also tracks just one dimension of scientific outputs. *Nature* has consistently advocated caution in the deployment of the impact factor in particular as a criterion of achievement (an index that Hirsch’s h indicator happily ignores). Wisely, the UK Research Assessment Exercise (RAE) has long committed itself to a broader view and the organizers of the next RAE, to take place in 2008, have

prohibited assessment panels from judging papers by the impact factors of the journals in which they appeared. What the costs of that will be in panel members’ time remains to be seen.

The common approach of the RAE’s disciplinary panels is to assess up to four submitted outputs (typically research papers or patents) per researcher, of which a proportion will be assessed in some detail (25% for the biologists, 50% for the physicists). There will no doubt be something of a challenge in taking into account the fact that a typical publication has several co-authors.

These outputs will sit alongside indicators of the research environment such as funds and infrastructure, and of esteem, such as personal awards and prestige lectures. The specific indicators to be considered and the weightings applied are now open for public consultation (see www.rae.ac.uk/pubs/2005/04/docs/consult.doc). Given that the RAE is so influential both nationally and, as a technique, internationally, there is a lot at stake. Stakeholders should express any concerns they may have by the deadline of 19 September. ■

“Given the potential for such indicators to be seized upon by administrators, readers should examine the suggestion and provide the author with some peer assessment.”

Climate for progress

The painstaking US approach to the assessment of climate-change science yields some useful results.

The US Climate Change Science Program, which is seeking to produce a comprehensive set of reports on climate change, has been widely criticized as a stalling exercise whose work duplicates that of the Intergovernmental Panel on Climate Change, and whose findings may be prone to political manipulation.

However, the programme yielded some promising outcomes last week, when several participants of a panel it convened published some findings on the apparent discrepancy between the record of warming on the Earth’s surface and the change experienced in the troposphere, the lowest level of the atmosphere (see page 896).

The climate-change programme had brought together 22 experts to look into the problem. In the course of five meetings and countless e-mail exchanges, the participants conceded flaws in some of their earlier analyses and produced new data sets, some of which were published online in *Science* (doi:10.1126/science1114772; doi:10.1126/science1114867; doi:10.1126/science1115640; 2005).

The release of the draft report summarizing the panel’s conclusions for public review, which was due in June, is being delayed, however. The delay may reflect the intrinsic difficulty of getting a large committee containing disparate views to agree on a final form of wording. But it seems to fit in with a pattern at the climate-change programme, which was criticized by the Government Accountability Office in April for failing to meet a 2004 deadline for issuing a new national climate-change assessment. Critics also point out that

since the programme issued a strategic plan just two years ago, some of the completion dates for the 21 reports that it plans to produce have slipped by several years.

US government officials say that they were too ambitious in setting the original deadlines. But they also admit that some of the panels’ work is being delayed by two other factors: the need to comply with recent legislation that makes it extremely difficult to incorporate fresh scientific information into government reports, and the requirement that the climate-change work is approved at the political, as well as the scientific, level.

The programme’s director, James Mahoney, argues that the latter process will be managed in a way that will preclude political interference. Once the scientific authors have agreed their final draft, the report will be posted online for a 45-day period of public comment. Once that closes, the report will be revised again. Only then, in the final step before publication, will it be reviewed by administration officials.

This process will make it hard for the Bush administration to put its own spin on the report’s scientific findings — provided that the drafts are published promptly, without political interference, and that the final documents adhere to the spirit of the drafts. The handling of the lower-atmosphere report — the first from the programme’s strategic plan — in the next few months will help clarify whether the programme is a serious attempt to grapple with uncertainty in climate-change science, or is merely an exercise in obfuscation. ■

“The handling of the lower-atmosphere report over the next few months will help to clarify whether the programme really is a serious attempt to grapple with uncertainty in climate-change science.”

RESEARCH HIGHLIGHTS

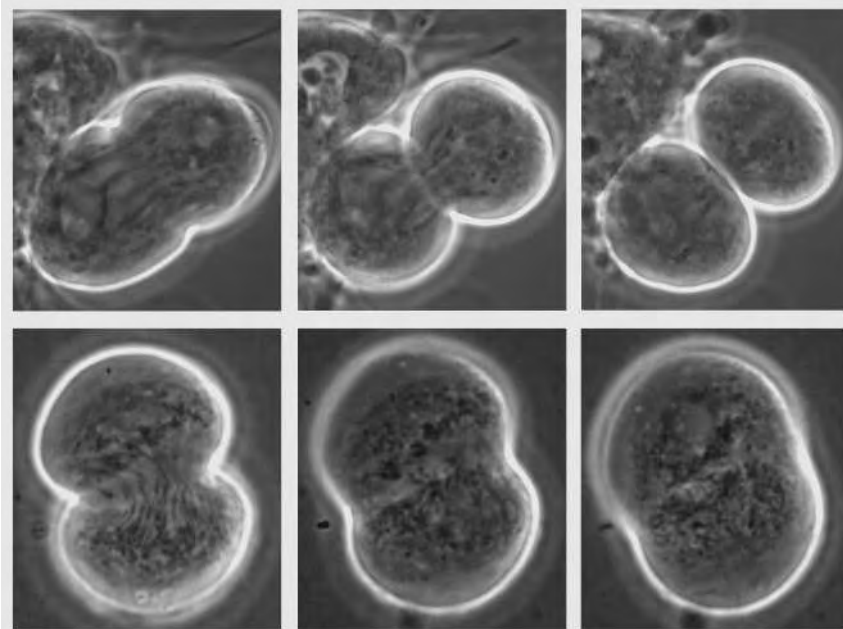
Getting good cleavage

Curr. Biol. **15**, 1401–1406 and 1407–1412 (2005)

The composition of the cell membrane and signalling via calcium ions are key to controlling the final step of cell division, according to two independent studies.

Julie Brill and her colleagues at the Hospital for Sick Children in Toronto, Canada, found an important component of the calcium signalling pathway, the phosphoinositide PIP2, in the dividing cell's membrane. Sperm cells from the fruitfly *Drosophila* deprived of either calcium or PIP2 (pictured bottom right, compared with control, top right) can begin cytokinesis, but not complete it, the researchers found.

Meanwhile, a group of researchers led by Seth Field of Harvard Medical School in Boston has shown that PIP2 accumulates at the cleavage furrow in mammalian cells, where the molecule helps the cell membrane to stick to the contracting ring of proteins that divides the cell.



R. WONG/J. BRILL/ELSEVIER

BIOTECHNOLOGY

Made to order

Nature Biotechnol. doi:10.1038/nbt1128 (2005)

Escherichia coli bacteria have been engineered to make novel polyketides — molecules that can kill bacteria and inhibit cancer cells — using a modular approach.

Typically, soil microorganisms make polyketides using hefty enzymes encoded in genetic sequences up to 50,000 bases long. To make these genes more wieldy, researchers headed by Daniel Santi from Kosan Biosciences in Hayward, California, divided the sequences into modules about 5,000 base pairs long. They then engineered modules from different enzymes to combine inside *E. coli*, allowing the bacterium to synthesize polyketides not found naturally. Such compounds could help fight drug-resistant pathogens.

PROSTATE CANCER

RNA interference delivers

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0501753102 (2005)

Small interfering RNA molecules have been shown to shut down disease-causing genes in many animal models, but an obstacle has been delivering the RNA to where it is needed. Takahiro Ochiya at the National Cancer Center Research Institute in Tokyo and his colleagues have pioneered one possible solution by combining small

interfering RNAs with a modified form of the natural protein collagen.

The group used such a mixture, injected into mice's bloodstreams, to shut down the genes implicated in the spread of prostate cancer to bone. This inhibited the cancer's spread, without the side-effects or inflammation that have plagued other approaches — providing the first hints that small RNAs could treat advanced prostate cancer.

COMPUTER SIMULATION

Cracked it

Phys. Rev. Lett. **95**, 060202 (2005)

Warming news for materials scientists: a way to simulate the behaviour of solids that previously worked only when the temperature of the material was set to absolute zero has been adapted to work at higher, more realistic temperatures.

Laurent Dupuy of the Lawrence Livermore National Laboratory in California and his colleagues have made this improvement to the quasicontinuum method, which combines atomic-scale simulations of crucial regions with cruder treatment of the rest. This technique speeds the simulation of phenomena such as fracture, which involve processes at scales from the atomic to the macroscopic. The team showcases its work by simulating the formation of dislocations in a flat nickel crystal at different temperatures.

METABOLISM

Melting fat

Dev. Cell **9**, 271–281 (2005)

The insulin signalling pathway controls whether the body burns or stores fat. Stephen Cohen's group at the European Molecular Biology Laboratory in Heidelberg, Germany, have now found that a protein called Melted is involved in regulating this pathway.

The researchers found that Melted, which resides in the cell membrane, binds to proteins from two different branches of the insulin pathway. Mutating the gene for Melted in the fruitfly *Drosophila* reduced the activity of the TOR pathway but activated FOXO, producing flies with 40% less body fat than normal. Giving the mutants the human version of the gene restored their fat, suggesting that Melted is involved in human fat metabolism, too.

ELECTRONICS

Y NOT?

Nature Mater.

doi:10.1038/nmat1450 (2005)

The first electrical switch made entirely from carbon nanotubes has been unveiled by Prabhakar Bandaru of the University of California, San Diego, and his colleagues. They report that a current

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

flowing across two branches of a Y-shaped nanotube can be switched on and off by applying a voltage to the third branch. Although the details of the switching process are not yet well understood, it should be possible to use the Y-shaped nanotube to create a NOT gate, which inverts its input, and to implement other logic operations.

ANTIBIOTIC RESISTANCE

TB's second secret

Proc. Natl Acad. Sci. USA

doi:10.1073/pnas.0505446102 (2005)

One reason for the spread of tuberculosis is the remarkable ability of its causative agent, *Mycobacterium tuberculosis*, to resist antibiotics. This stems, in part, from the low permeability of the cell envelope, a membrane inside the organism's cell wall. Now a gene that coordinates a second line of defence within the cell's cytoplasm has been discovered.

Charles Thompson from the University of British Columbia in Vancouver and his team spotted that a mutation in the gene *whiB7* made the bacterium *Streptomyces lividans* more vulnerable to antibiotics. *Mycobacterium* also carries this gene, and mutating it had the same effect. The researchers analysed gene activity to study the protection mechanism.

QUANTUM PHYSICS

It must be glove

Phys. Rev. A **72**, 022304 (2005).

Information about handedness can be transmitted by a physical object such as a glove, because the left hand is different from the right. But what would a pair of 'quantum gloves' look like?

Daniel Collins, formerly of the University of Geneva, Switzerland, and his colleagues point out that four entangled particles can assume one of two configurations that are, like gloves, mirror images of each other.

But these gloves are much more economical than a classical glove, they argue, because they are burdened with less of the extra information that comes from knowing the position of every part of a classical glove with certainty.

In contrast, the absolute positions of the particles in

their quantum glove are unknown, so they only carry information about handedness.

CELL BIOLOGY

Nuclear escape

Cell **122**, 379–391 (2005)

The editing of messenger RNA, or splicing, was assumed to occur only in a cell's nucleus. But now splicing has been seen in human platelets, which have no nucleus.

Mature mRNA molecules, which are used to synthesize proteins, are made by splicing of pre-mRNA. Andrew Weyrich's group at the University of Utah in Salt Lake City identified components of the splicing apparatus in platelets (pictured below, surrounded by red blood cells), and showed that external signals

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

D. SCHARE/SPL

could set the machinery going. This prompted them to examine bone-marrow cells called megakaryocytes, which spawn platelets. Parts of the splicing complex were also found in these cells' cytoplasm.

GENETICS

Surprise connection

Mol. Cell **19**, 381–391 (2005)

Inside the cell nucleus, DNA wraps around proteins called histones. Chemical modifications of these proteins are thought to switch genes on and off. A study of the histone H3 provides clues to how sophisticated this control may be.

Adding methyl groups to the ninth amino acid in this protein's chain had been thought to deactivate genes. Gerd Blobel of the Children's Hospital of Philadelphia and his colleagues were therefore surprised to find di- and tri-methylation of this region in active genes. These modifications disappear when the gene is repressed. Other histone modifications must also be influencing the gene's status.

DORLING KINDERSLEY/GETTY IMAGES

JOURNAL CLUB

Douglas Hamilton
University of Maryland

Now you see them, now you don't, but at least this researcher knows why his favourite features of Saturn's rings have gone missing.

Images from the Voyager spacecraft in the 1980s revealed dark streaks across Saturn's rings, dubbed spokes. They grew rapidly to span thousands of kilometres, sheared apart, and then faded over a few hours.

These unusual structures are known to be composed of micrometre-sized dust grains, made by meteoroid impacts on large ring particles, but their evolution is not well understood. I had been working on this problem, and had come up with some very promising leads, based on models of the motions of charged dust grains. However, the late 1990s brought data that threatened to make my theory redundant.

Colleen McGhee from Wellesley College in Massachusetts and her colleagues were using the Hubble Space Telescope to monitor Saturn's rings. They found, mysteriously, that spokes lessened in number and intensity over several years, then disappeared altogether in late 1998. I learnt about these observations at meetings, and was left wondering what had happened to the spokes.

McGhee *et al.* now offer an elegant explanation for the spokes' remarkable disappearing act (*Icarus* **173**, 508–521; 2005). They suggest the spokes form a dusty scattering layer above the thin main rings. Until the mid-1990s, Saturn's rings were seen edge-on, so sunlight had to travel a long distance through the dusty layer before reflecting from the rings, making the spokes look dark.

Now, Saturn's rings are tilted towards the Sun. This geometry makes the spokes invisible to Earth-based telescopes, and even to the Cassini spacecraft currently orbiting Saturn. But the dusty spokes still exist, and so my budding theory survives.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

NEWS

Synthetic biologists face up to security issues

WASHINGTON DC

Concerns about safety and security in the young field of synthetic biology are developing almost as rapidly as the research itself.

With scientists now able to create complete genomes from scratch and to introduce new characteristics into viruses and bacteria, there are fears that accidental — or worse, intentional — release of such creatures could occur.

As a result, the US government, academics and the field's first private companies are now working together to address the potential problems before they become a reality.

Synthetic biology is one of the priorities for the newly formed US National Science Advisory Board for Biosecurity (NSABB), which met for the first time in June. A working group on the topic, set up by the board, will begin its discussions later this month. Its members have already consulted synthetic biologists who have set up their own governance project, funded by the Alfred P. Sloan Foundation in New York, which will have its first meeting in September.

And one of the first public discussions of the issue comes this week, at a synthetic-biology meeting in San Francisco on 19–20 August. As well as looking at the field's potential in areas such as drug development, cellular reprogramming and biological robotics, participants will tackle ethical and legal issues.

The discussion will echo an event in 1975, when pioneers of genetic engineering retreated to the Asilomar Conference Center in California to set safety and ethical principles to guide their field. The fear was that modified microorganisms might escape into the environment with unpredictable effects, perhaps causing disease or out-competing wild strains.

Now synthetic biologists have much more powerful techniques at their disposal. Improved DNA synthesis means, for example, that microbial genomes can be built from scratch, bypassing the need to get hold of the actual organism. A dramatic demonstration of the potential consequences came in July 2002, when researchers at the University of New York at Stony Brook reported that they had synthesized an infectious poliovirus using mail-order DNA (see *Nature* **418**, 265; 2005). Genome sequencer Craig Venter followed that by announcing plans to build a bacterium from lab synthesized DNA and by taking just three weeks to synthesize a virus that infects

bacteria (H. O. Smith *et al.* *Proc. Natl Acad. Sci. USA* **100**, 15440–15445; 2003).

Others are using made-to-order components to re-engineer the genomes of bacteria or viruses — either to investigate how they work or to try to give them abilities that they would not have naturally, such as producing drug molecules or generating hydrogen for use as an energy source. In theory, these techniques could be used to create deadly organisms in the lab, such as Ebola or anthrax, or to make bacteria more dangerous by giving them antibiotic resistance, for example, or the ability to make additional toxins.

Does this increased sophistication mean that the field needs radically new rules? Drew Endy, a synthetic biologist at Massachusetts Institute of Technology and one of the leaders of the Sloan Foundation project, is not sure. “There hasn't been a clear conversation about whether *de novo* synthesis poses a different threat from genetic engineering,” he says.

Wendell Lim at the University of California, San Francisco, who uses synthetic biology to study basic processes such as how cells grow and move, says that there is nothing special about the field that requires blanket restrictions and that each type of study should be judged on its own risks and merits. “To broadly raise alarms about all such approaches is akin to saying we should worry about all uses of semiconductors because one of their uses could be in launch systems for nuclear weapons,” he says. “We need to distinguish the sub-areas of the field and specific biological components that truly pose a hazard, and figure out how to regulate them.”

Taking care

But there are some general precautions that most synthetic biologists support. Many think it would make sense to monitor products ordered from companies that generate synthetic genes on demand, such as Codon Devices in Cambridge, Massachusetts, and Blue Heron Biotechnology in Bothell, Washington.

George Church, who co-founded Codon Devices this spring, wants to go further, arguing that strict regulation is a key to heading off the public-relations problems that have plagued other areas of biotechnology. As well as monitoring orders for components such as synthetic genes, he says the government

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The ability to build genomes in the lab has raised fears that the technology could be abused.

should screen certain raw materials that scientists could use to brew up their own DNA. “Basically you want a series of licences to cover almost every step where there's a reasonable bottleneck you can regulate,” he says.

But many argue that the openness of the field offers a built-in safety mechanism against unintentionally creating something harmful. For instance, Endy's lab hosts an online library of parts that can be built into genomes. The library is open source, meaning that all scientists in the field can test them. “There's an element of safety in using well characterized components,” says Christopher Voigt, a synthetic biologist at the University of California, San Francisco.

Other synthetic biologists point out the success of initiatives to encourage young scientists to think about safety and ethical issues relating to their work. “It's important to talk about ways to minimize risk, and one of the most important ways to do that is to train responsible scientists and engineers,” says Christina Smolke, a chemical engineer at the California Institute of Technology in Pasadena.

**MARS SOCIETY**

Follow the action from the meeting of Mars enthusiasts in Boulder, Colorado.

www.nature.com/news

S. OGDEN/SPL

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

NASA draws up blueprint for craft to reach Moon and Mars

BOULDER, COLORADO

NASA last week revealed its plans for the next generation of space vehicles, designed to get humans back to the Moon and, eventually, to Mars.

On 13 August, Christopher Shank, special assistant to NASA administrator Michael Griffin, described the agency's future exploration machinery, which is proposed to replace the shuttle in 2010.

His talk, given at the Mars Society Conference in Boulder, Colorado, precedes a full report scheduled for release this month, called the Exploration Systems Architecture Study. This will lay out how NASA intends to meet President George W. Bush's goal of sending humans back to the Moon by 2020 in preparation for a Mars mission.

"It will be a 'go as you can afford to pay' approach," Shank told the 300 members of the Mars Society, a private group whose mission is to promote human exploration and colonization of the red planet. "Let the long, hard slog begin."

That, he said, means deferring other programmes, such as future research on the International Space Station (ISS) and lunar-base development, until the new space vehicles come online. The first vehicle will be a 25-tonne Crew Exploration Vehicle (CEV), which will ride into space on a modified shuttle booster rocket; this will initially supply and return ISS crews. After 2010, work will begin on a 100-tonne Heavy Lift Vehicle (HLV) that can carry heavier payloads. The HLV will also use a rocket modified from the shuttle's boosters and external fuel tank.

Lockheed Martin and a team at Northrop Grumman-Boeing have contracts to develop the three- to six-person CEVs. Potential designs include

a large Apollo-like capsule or a smaller, slimmer shuttle. NASA will pick the winning design in 2006.

Mars Society members were thrilled to hear about the vehicle plans and they support Griffin in getting started. But they object to the proposed timeline, and the size of the CEV. They want to see a 7-tonne vehicle being used as this could take humans directly to the Moon and back, and so accelerate lunar-base construction. NASA's 25-tonne CEV could carry six people as well as ISS supplies and parts. But to reach the Moon it would need to

rendezvous with another, as yet undeveloped, vehicle in lunar orbit before returning to Earth.

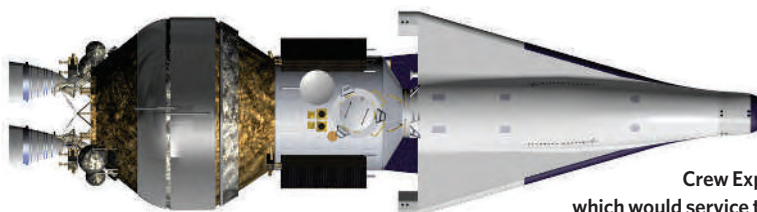
"It's a choice between having a Cadillac ISS programme or a lunar

base," says Robert Zubrin, society president. He argues that continuing to focus on the ISS does little to further the goal of getting to Mars. Shank counters that NASA does not separate the ISS from Moon and Mars missions, as the same components are integral to both.

Zubrin and his fellow enthusiasts applauded the plans for an HLV, which is critical for a martian journey. But some, such as David Schuman, a lawyer at NASA's Goddard Space Flight Center in Greenbelt, Maryland, criticized the decision to delay HLV development until after 2010.

Without solutions to the shuttle's current safety issues, he notes, "we could be left without any heavy-lift capability". That would delay ISS completion, pushing Moon and Mars timelines back another decade. As it is 36 years since the first lunar landing, Zubrin and fellow exploration proponents say that delay is not only unbelievable — it is unacceptable. ■

Kendall Powell



One of the designs for the Crew Exploration Vehicle, which would service the space station.

LOCKHEED MARTIN

So far, observers are giving synthetic biologists credit for tackling such issues. But some, such as Paul Rabinow, an anthropologist at the University of California, Berkeley, think that the field needs to take a broader view than it has so far. He is speaking at the San Francisco meeting this week and says he will tell attendees to pay more attention to the danger that synthetic biology could be abused for biowarfare or bioterrorism, something that is harder to control than the risk of accidents.

"The field has been attempting to turn this into a safety issue, but we're living today in a security regime," he told *Nature*. "Dealing with safety is good, but they're fooling themselves if they think that's going to be the end of this question."

Venter agrees, arguing that rather than regulating labs and companies, governments need to use synthetic biology to develop ways to defeat bioterrorist attacks, such as producing drugs and vaccines, or sensors to detect altered organisms in the environment. "If we're not concentrating 100% of our defensive effort on countermeasures, I think we're missing the big picture," he told the NSABB in June. ■

Erika Check

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Warming debate highlights poor data

An unusual truce has been reached in the turbulent field of climate science. Scientists who have spent 15 years arguing over a discrepancy in certain data on global warming now say they all agree: the data are inadequate.

The lowest layer of Earth's atmosphere is indeed warming, but uncertainties in the data are as large as the trends the scientists are looking for, says Peter Thorne, a researcher with the UK Met Office. "The fact that the uncertainty has increased is actually a step forward," he says.

Thorne is one of the lead authors on a new report commissioned by the US Climate Change Science Program (CCSP), the government entity responsible for research on climate change. The report will be released for public review in the next few months, but three papers published online last week by *Science* preview some of its findings.

The papers advance a debate that began in 1990, when an analysis of satellite observations suggested that temperature trends in the troposphere, the lowest layer of the atmosphere, were inconsistent with the picture of global warming emerging from measurements made at the surface¹. This finding became ammunition for sceptics arguing against evidence for global warming.

But the three new papers add important new results, says Carl Mears of Remote Sensing Systems in Santa Rosa, California, and a co-author of one of the reports. "We are converging," he says. "We are definitely getting closer."

Along with colleague Frank Wentz, Mears reanalysed raw satellite data, including those used in the original 1990 study. Their analysis² suggests that the troposphere is warming at a rate of 0.19 °C per decade — far faster than the estimate of 0.09 °C reported by a group at the University of Alabama in Huntsville, who did the original study.

But recently the Alabama team issued its own revised data set, reporting a warming of 0.12 °C per decade. The group has adopted a new way to adjust the satellite data for the time of day the observations were taken. The change comes in response to a problem pointed out during a meeting of authors for the forthcoming CCSP report.

"What we all show is that the troposphere is warming. The question is by how much," says climatologist John Christy of the Alabama team. "When we are talking about precisions of just a few hundredths of a degree per decade, we're not quite there with our observing systems."

The second *Science* paper points out problems in the temperature record taken by weather balloons, in part because different manufacturers' instruments heat up by different amounts during the day³. The third paper reviews predictions from 19 different climate models and concludes that the differences between the models' predictions and observations are most likely to be the result of errors in the observations and in how they had been analysed⁴.

Together, the reports conclude that the observed tropospheric temperature trends are consistent with a warming world. Yet climate experts acknowledge that the papers are just another data set to argue over, as happened with an earlier analysis published in *Nature*⁵ that failed to lay the argument to rest.

"I don't think this will be the last word," says climate researcher Phil Jones from the University of East Anglia.

Jenny Hogan

1. Spencer, R. W. & Christy, J. R. *Science* **247**, 1558–1662 (1990).
2. Mears, C. A. & Wentz, F. J. *Science* doi:10.1126/science.1114772 (2005).
3. Sherwood, S., Lazante, J. & Meyer, C. *Science* doi:10.1126/science.1115640 (2005).
4. Santer, B. D. et al. *Science* doi:10.1126/science.1114867 (2005).
5. Fu, Q., Johanson, C. M., Warren, S. G. & Seidel, D. J. *Nature* **429**, 55–58 (2004).

Climate sceptics place bets on world cooling down

A British climate modeller has finally persuaded global-warming sceptics to wager money on their contrarian predictions about climate change.

James Annan, who is based at the Japan Agency for Marine-Earth Science and Technology in Yokohama, has agreed a US\$10,000 bet with Galina Mashnich and Vladimir Bashkirtsev, two solar physicists who argue that global temperatures are driven by changes in the Sun's activity and will fall over the next decade. The bet, which both sides say they are willing to formalize in a legal document, came after other climate sceptics refused to wager money.

Annan began his quest last winter after hearing Richard Lindzen, a meteorologist at the Massachusetts Institute of Technology who questions the extent to which human activities are influencing climate, say he was willing to bet that global temperatures will drop over the next 20 years. "A pay-off at retirement age would be a nice top-up to my pension," says Annan.

But no wager was ever agreed. Annan says that Lindzen wanted odds of 50-to-1 against falling temperatures: this meant that Annan would pay out \$10,000 if temperatures dropped, but receive only \$200 if they rose. In total, Annan says he tried and failed to agree terms with seven sceptics.

Other potential climate gamblers have drawn a blank with their attempts to enter similar bets with climate-change sceptics. In May, environmental activist George Monbiot challenged climate sceptic Myron Ebell to a £5,000 (US\$9,000) wager live on

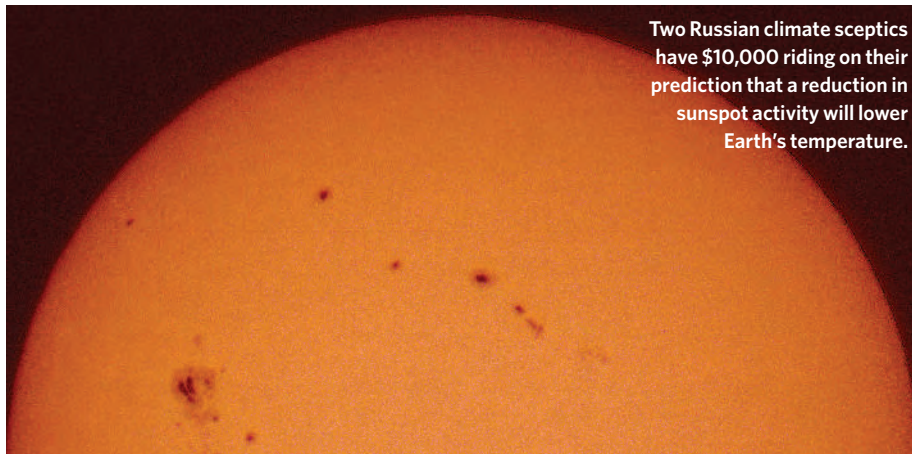
BBC radio. Ebell, a global-warming specialist at the Competitive Enterprise Institute, a think-tank in Washington DC, declined, saying he has four children to put through university and so does not "want to take risks".

But Annan's search ended with Mashnich and Bashkirtsev, who are based at the Institute of Solar-Terrestrial Physics in Irkutsk, Russia. They say that global surface air temperatures closely correlate with the size and number of sunspots. Sunspot levels follow regular patterns and the Sun is expected to be in a less active phase over the next few decades, leading Mashnich and Bashkirtsev to predict a drop in temperature.

Both sides have agreed to compare the average global surface temperature between 1998 and 2003 with that between 2012 and 2017, as defined by the records of the US National Climatic Data Center. If the temperature drops, Annan will pay Mashnich and Bashkirtsev \$10,000 in 2018, with the same sum going the other way if the temperature rises.

Piers Corbyn, head of Weather Action, a private meteorological service based in London, told *Nature* he would like to enter into a similar bet. Corbyn's theory, the details of which he has not revealed, predicts that changes in solar activity will cause "considerable world cooling" by 2040. Annan challenged him to a bet in May, but Corbyn says he did not receive the e-mail. "I'm happy to bet loads of money," he says. ■
Jim Giles

Two Russian climate sceptics have \$10,000 riding on their prediction that a reduction in sunspot activity will lower Earth's temperature.



D. MILLER/DMI

ON THE RECORD

“It is not enough to be a corpse any more. Now you have to be a politically correct corpse.”

Michigan funeral director Thomas Lynch on ‘green’ cemeteries, which use biodegradable coffins.

“It could potentially be a showstopper if somebody gets suicidal on the way to Mars.”

Psychiatrist Nick Kanas tells NASA it should worry about astronaut stress.

“We realized that using a cruise liner makes it impossible for many participants to rightfully cover their expenses.”

A meeting organizer explains why the First International Symposium on Nanotoxicology has shifted venue from a Caribbean cruise ship.

SCORECARD**White bread**

After eight years and millions of dollars spent on research, US scientists figure out a way to sneak whole grains into white loaves.

**Urine**

Singaporean physicists have developed a paper battery that generates power from pee. They say it could be used in cheap biological home-testing kits, but will consumers go for it?

**Lab safety**

An employee at Los Alamos National Laboratory accidentally brings americium-241 to family homes, and ships some to Pennsylvania, triggering a four-state sweep for radiological material.

DATAPOINT

Percentage of scientists who say they don't believe in God.

Biologists	41%
Physicists	41%
Chemists	27%

Source: Elaine Ecklund and Christopher Scheitle, unpublished work presented at the 2005 meeting of the Association for the Sociology of Religion.

Survey questions safety of alternative medicine

Practitioners of complementary and alternative medicine (CAM) are not doing enough to protect their patients, according to an expert in the field.

A survey of 95 British CAM organizations revealed that few practitioners monitor the side effects of their therapies. And many of the organizations questioned said that they don't even consider the possibility that adverse reactions might occur.

Edzard Ernst, professor of complementary medicine at the Peninsula Medical School of the Universities of Exeter and Plymouth, wrote to the CAM organizations in May, asking whether they advise their members to report adverse events. He also asked for details of such events. The organizations he approached covered some 20 healing arts, from herbal medicines and homeopathy to crystal therapy and hypnotherapy.

Fewer than a third replied. Of those who did, just nine said that they advised members to report side effects. And only one, an acupuncture association, gave details of adverse reactions reported in 2004.

“Several organizations said that adverse events were only connected with mainstream medicine, but were inconceivable in their own practice,” Ernst says.

Side effects do occur with CAM and can sometimes be serious — stroke can occur after chiropractic treatments and some deaths have been recorded. “Yet no one is looking,” says Ernst. “In the interest of patient safety, that has to change.” He wants studies to be carried out in different parts of the world to research adverse events associated with each therapy.

“We have been calling for more regulation for a long time,” says a spokesperson for the British Medical Association. “Doctors know that patients like these sorts of therapies, and they would like to be confident before they refer a patient to a CAM practitioner that the therapy in question works and is safe.”

Terry Cullen, chairman of the British Complementary Medicine Association, says that his organization does not believe the therapies it represents could be harmful. For example, he says that even if a client receiving reflexology, which involves applying pressure

to the hands and feet, doesn't end up more relaxed, they are unlikely to have any reaction serious enough to need monitoring or regulating.

But the Scottish Institute of Reflexology says that Ernst's enquiry has prompted it to establish a system for monitoring adverse effects.

“We are going to introduce a sort of ‘yellow-card’ system like the medical profession has — a form that members can fill in and return to us if they observe any adverse effects,” says Margaret Smillie, the institute's president. Smillie adds that comple-

mentary medicine should adopt the same standards as the mainstream. “We feel a responsibility to protect the public and our members,” she says.

Alison Abbott

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Little is known about possible side effects of complementary treatments.

Ernst says that he was surprised by the general lack of familiarity with the concept of adverse reactions. “Some respondents said they didn't really understand what was meant by the term,” he says. Others denied its rele-


**ECOLOGICAL SOCIETY
OF AMERICA**

Catch up on this year's meeting in Montreal, with news stories and a blog.
www.nature.com/news

A. FAVILA/AP

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

WHO urges regional offices to stockpile flu drug for staff

Has your employer thought about how to protect its staff and maintain essential services in the event of a global influenza pandemic? If not, it's worth noting how seriously the organization perhaps best informed about the threat — the World Health Organization (WHO) — is taking the possibility. Its own plan calls for a stockpile of antiviral drugs large enough to treat up to 30% of its staff and their families.

Experts agree that the H5N1 avian flu virus,

which is endemic in southeast Asia and is fanning out across Russia and Kazakhstan, could trigger a human pandemic. The WHO's internal plan, a copy of which has been obtained by *Nature*, warns that all prerequisites for the start of a pandemic have been met except one: the establishment of efficient and sustained human-to-human transmission of the virus.

The document, dated 30 May 2005, notes that if a pandemic occurs, the WHO has a duty

Shigeru Omi, WHO regional director for the western Pacific, plots the spread of avian flu.

of care to its staff and their dependents, and to maintain the agency's essential functions. It adds that if a pandemic is declared, "it is very likely that all stocks of medicine useful against influenza, particularly oseltamivir, will be rapidly exhausted". It suggests that WHO offices stockpile the drug for 8,000 employees.

Each WHO office should have enough oseltamivir (Tamiflu) to give a five-day course to 30% of its staff and their dependents, the plan suggests. "Some countries may choose to stockpile more, but should be aware that it may become difficult to reserve these excess stores for only the use of UN personnel and families during a severe community-wide outbreak." It also advises that, "because antivirals will become valuable commodities during a pandemic, they should be stored in a secure place".

According to the document, the WHO's headquarters in Geneva is initially stockpiling 1,000 five-day courses of oseltamivir. The plan also lists a series of secondary measures such as stockpiling antibiotics, syringes and face masks, and instructs offices to consider how to convert warehouses, meeting rooms and gymnasiums into temporary wards.

The United States is expected to release its national pandemic plan in the next few weeks. France and Britain have ordered drug stockpiles to cover a quarter of their citizens, close to WHO's internal target, but the US stockpile currently covers just 1% of its population. ■

Declan Butler

Kansas backs lessons critical of evolution

The Kansas State Board of Education has decided by six votes to four to include stronger criticism of evolution in its high-school biology curriculum. Science advocates fear that the move paves the way for 'intelligent design' — the idea that an intelligent creator shaped living things — to reach the classroom.

"This is a religiously motivated strategy," says Harry McDonald, president of Kansas Citizens for Science in Olathe, which vehemently opposes the new standards. "Religious advocates have decided that they can push their views forward by casting doubt on science."

The Kansas Board of Education's chairman Steve Abrams, who helped write the amendments, dismisses the charges as "baloney". "Is it wrong to teach critical

analysis and critical thinking?" he asks.

This is the second time that the board has tried to alter the state's standards for the teaching of science. In 1999, the board voted to remove evolution, cosmology and geology from Kansas's curriculum, leaving teachers to decide individually whether or how to teach the subjects. But these topics were reinstated into the high-school syllabus after activists from science and business communities helped elect more moderate members to the board in 2000 (see *Nature* 406, 552; 2000).

Now, a newly elected, more conservative board is seeking to augment the evolution curriculum with criticisms of Darwin's theory commonly espoused by the intelligent-design movement. The criticisms include gaps in the fossil record and the inability of evolution to explain the first life

on Earth. The standards also refer to macroevolution as a "controversial" theory.

"These standards are very clearly denigrating evolution," says Eugenie Scott, executive director of the National Center for Science Education in Oakland, California. She believes that the standards are part of a new nationwide strategy by intelligent-design advocates to undermine the way evolution is taught in public schools. In 2002, Ohio passed education standards that mandate teaching that scientists "continue to investigate and critically analyse aspects of evolutionary theory". But the new Kansas rules are far bolder, says Scott.

The board will now send the new standards out for an external review, with a final vote scheduled for later this autumn. ■
Geoff Brumfield

Index aims for fair ranking of scientists



This year's intake at the National Academy of Sciences matched the anticipated ranking on the h-index.

The election procedures of scientific academies are often seen as opaque, clubby and capricious. But Jorge Hirsch, a physicist at the University of California, San Diego, may have found a way to silence those complaints, by inventing a measure of research achievement that, he says, is transparent, unbiased and very hard to rig.

His 'h-index' depends on both the number of a scientist's publications, and their impact on his or her peers. As well as determining membership of scientific societies, Hirsch suggests that the method could inform funding or tenure decisions.

"It's a very cute idea," says Sidney Redner, a physicist at Boston University who has studied scientific citation statistics. He welcomes an alternative to simplistic readings of such statistics, which "leave so much room for misinterpretation".

Redner also agrees that it would be useful to have an objective criterion for election to bodies such as the US National Academy of Sciences (NAS) or Britain's Royal Society. He has been involved in choosing fellows of the American Physical Society (APS), and says that factors other than research quality inevitably come into play. "If you're not a political person, you don't get nominated," he says. A systematic criterion such as the h-index "would make the playing field more level".

The h-index is the highest number of papers a scientist has that have each received at least that number of citations. Thus, someone with

an h-index of 50 has written 50 papers that have each had at least 50 citations. This, says Hirsch, is fairer than alternative measures based on publication. Counting total papers, for example, could reward those with many mediocre publications, whereas just counting highest-ranked papers may not recognize a large and consistent body of work.

Some of the highest-ranked physicists, by h-index



110 Ed Witten (pictured)
Princeton Institute for Advanced Study. Devised M theory.

94 Marvin Cohen University of California, Berkeley. Condensed-matter theorist.

91 Philip Anderson Princeton University. Condensed-matter theorist, won Nobel prize in 1977.

86 Manuel Cardona Max Planck Institute for Solid State Research. Works on superconductors.

79 Pierre-Gilles de Gennes ESPCI, Paris. Condensed-matter theorist, won Nobel prize in 1991.

68 Frank Wilczek Massachusetts Institute of Technology. Won Nobel prize in 2004 for work on the strong force.

66 David Gross Kavli Institute for Theoretical Physics, Santa Barbara. Won 2004 Nobel prize with Wilczek.

And it is hard to inflate one's own h-index, for example by self-citation. "You can't fake it," says Hirsch, because it relies on how a body of work is received over time. "To manipulate an entire career is very hard," agrees Redner.

Applying the method to physicists certainly seems to pick out influential individuals (see Box). Top, by a considerable margin, is Ed Witten of the Princeton Institute for Advanced Study, with an h of 110. Witten, who devised the extension of string theory known as M theory, is widely regarded by his peers as the most brilliant living physicist.

Hirsch suggests that after 20 years in research, an h of 20 is a sign of success, and one of 40 indicates "outstanding scientists likely to be found only at the major research laboratories". An h of about 12 should be good enough to secure university tenure, he says, and fellowship of the APS, for example, should occur typically for an h of 15–20, and of the NAS for an h of about 45. In 2005, new NAS members in physics and astronomy had an average h of 44.

Just desserts

Different disciplines have different citation patterns, says Hirsch, so each field would need different thresholds. Biologists can have h values of up to 190. But with that proviso, the method should work across disciplines.

One of the index's main attractions is that it can also rescue from obscurity researchers who have made sustained and significant contributions but who have not won the reputation they deserve. Many solid-state physicists would applaud the contributions of Manuel Cardona (h of 86) at the Max Planck Institute for Solid State Research in Stuttgart, Germany. But few might have ranked him alongside Nobel laureates Philip Anderson (h of 91) and Pierre-Gilles de Gennes (h of 79; see Box).

But the research community may take some convincing to put its faith in numbers rather than judgement. Ed Hughes, manager of the UK Research Assessment Exercise, which assesses the quality of university science departments to determine their funding, says that the exercise purposely avoids metrics in favour of expert review panels. "We explicitly don't use impact factors and citation indices," he says, explaining that 96% of researchers consulted after the 2001 assessment were in favour of using peer review.

Philip Ball

♦ <http://xxx.arxiv.org/abs/physics/0508025>

M. TEMCHINE

R. HAGADORN/INST. FOR ADVANCED STUDY

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Helping hand: African ecologists battling against desertification will get assistance from Britain.

UK ecologists pledge \$1m to keep in touch with Africa

The British Ecological Society has pledged US\$1 million over the next five years to support environmental researchers in developing nations, mostly in Africa.

The initiative is aimed at boosting communication between scientists in poor countries and established ecological societies in the developed world, says the society's president, Alastair Fitter. He made the announcement on 9 August in Montreal, at a gathering of the heads of 13 ecological societies. Participants emphasized the importance of helping local ecologists find regional solutions to environmental issues such as habitat loss, desertification and species conservation.

"The major environmental challenges faced by developing nations will require new scientific understanding as well as infrastructure and tools," says Jerry Melillo, outgoing president of the Ecological Society of America, which will hold a conference in Mexico in January to foster research by young ecologists in Latin America.

Leaking Japanese X-ray satellite loses its cool

Japan's X-ray astronomy satellite Suzaku has suffered a crippling loss. On 8 August, the Japan Aerospace Exploration Agency discovered that all the liquid helium inside the satellite's X-ray spectrometer had evaporated because of a leak.

Suzaku's scientific capability will be severely curtailed because the helium kept the instrument's sensitive detectors cold enough to function. The spectrometer was designed to measure X-ray intensities using detectors 20 times more sensitive than those on ordinary instruments. Astronomers had planned to use it to measure gas moving in galaxy clusters and around black holes.

"The loss of the X-ray spectrometer has a large impact on Suzaku and high-energy

astrophysics," says project scientist Richard Kelley of NASA. This is the second blow for Japan's X-ray astronomy programme. Suzaku was a replacement for Astro-E, which suffered a launch failure in 2000.

No news is good news for donations to tissue banks

Even positive media coverage does more harm than good to tissue banks, says an analysis of the media's impact on tissue donations in Britain.

The study, published on 13 August in the *BMJ*, catalogued UK newspaper stories between 1998 and 2004 on children's tissue being removed for research purposes. The stories included coverage of a scandal in 1999, when organs were found to have been removed from hundreds of dead children at a Liverpool hospital without their parents' consent. Two years later, there were fewer than 100 donations to tissue banks during a six-month period, compared with more than 200 over a comparable period before the scandal. Donations dropped as media coverage of tissue donation increased — even when newspapers ran positive stories.

The study's authors, led by Mary Dixon-Woods of the University of Leicester, think that media scrutiny makes families reluctant to donate tissue, and the staff at tissue banks feel squeamish about requesting such donations. "Media reporting of science can have important implications for those who conduct and regulate science," the team writes.

Multimillion-dollar physics experiment cancelled

The US National Science Foundation has cancelled a multimillion-dollar physics experiment over fears that the proposed project may run over budget.

All eyes on Madagascar's tiny discovery

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The Rare Symmetry Violating Processes (RSVP) experiment was to use an accelerator at Brookhaven National Laboratory in Upton, New York, to probe extremely rare particle decays in the hope of refining the standard model of particle physics.

The project was originally budgeted at US\$145 million, with construction slated to begin this year. But a 2004 analysis warned that project costs could double, in part because RSVP would require extensive upgrades to an existing accelerator. Based on that review, the National Science Board, which oversees the foundation, announced on 11 August that it was cancelling the RSVP project.

Laurence Littenberg, a Brookhaven physicist and co-spokesperson on one of RSVP's two experiments, says it could be a decade before another group performs similar work, perhaps at the Japan Proton Accelerator Research Complex that is currently under construction.

Biologist takes control of science at NASA

NASA has named former astronaut Mary Cleave as the head of its science office. She replaces Alphonso Diaz as associate administrator for science. The change was one of several announced last week during a shake-up by NASA chief Michael Griffin.

Cleave is a biologist and engineer who has served as project manager for the SeaWiFS satellite, which monitored marine chlorophyll on a global basis, and then headed NASA's Earth-Sun research division. Her deputy is Colleen Hartman, who worked as agency liaison with the Office of Science and Technology Policy, which advises the White House on science issues.

The appointments were announced on 12 August, although rumours had floated around Washington for several months.

A new species of mouse lemur, shown left, has been found thriving in a relatively well-studied area in the eastern rainforests of Madagascar. Genetic analyses revealed that the creature, dubbed *Microcebus lehilahytsara*, probably diverged from the other mouse lemurs that lived nearby more than two million years ago.

The same study, which was carried out by researchers at the German Primate Center and Georg-August University of Göttingen, also identified the second known species of giant mouse lemur, *Mirza zaza*, which is about the size of a grey squirrel. Together the discoveries bring the number of known lemur species to 49.



Spin doctor: Rhea Seddon (right) tests the biological effects of the space shuttle's rotating chair.

Testing times

Preparing astronauts for a journey to the red planet has become NASA's research priority for the International Space Station. But such experiments will need more than the skeleton crew now running the station. **Tony Reichhardt** reports.

At a meeting in Annapolis, Maryland, last February to draw up a 'roadmap' for NASA's use of the International Space Station, Terri Lomax, a research manager for the agency's exploration directorate, recounted the list of medical issues that worry doctors about sending astronauts on a 30-month round trip to Mars.

She put up a chart showing that 588 of 607 space-shuttle crew had reported 'medical events' or symptoms during their flight. The implication was that nearly everyone got sick. Astronaut Ed Lu, sitting towards the back of the room, raised his hand. "But aren't most of those stuffy noses?" he asked.

Lomax conceded as much, then presented statistics on the deterioration of bone in orbit — astronauts don't want brittle bones when they land on Mars. Lu raised his hand again. He spent six months on the space station in 2003, and gained, not lost, bone mass. So had

other recent space-station astronauts. The difference? They had a treadmill, so they could exercise to counter the debilitating effects of weightlessness.

Lomax then moved on to radiation, maybe the biggest threat of all. She said more dosimeters were needed on the station to characterize the risk. Up shot Lu's hand. We already know about the radiation environment in space, he said. What we don't know is how it would affect interplanetary travellers — and studies on the station would not help, because it is shielded from the most dangerous radiation.

The space station should be a unique place to do science, including the sort of biomedical testing needed before NASA sends astronauts to Mars. But some people question whether the right experiments are being planned, and

if so, whether the station will be equipped to answer them.

It's a key concern not just for astronauts, but for the \$30-billion space station, which is still only half finished 21 years after design work began. Europe and Japan plan a mixed portfolio of physics, biology and materials science for their station modules, which they hope to see launched by 2010 (see 'Stalled countdown', opposite) despite the space shuttle's current woes. But NASA, the project's main underwriter, has recently settled on human research as the top priority for the orbiting laboratory.

Ever since 2004, when the White House announced plans to send people to the Moon and on to Mars, NASA has been reshaping its space-station research to support the new goal. Details of the Moon-Mars programme are expected at the end of this month. Some

in Congress want the agency to keep a percentage of station research for fundamental experiments in biology, physics and materials science. But biomedical studies are clearly the new focus.

Speaking to a congressional committee in June, NASA's new administrator, Michael Griffin, listed the agency's priorities for space-station science: all involved either medical research on humans or work on life-support systems to keep people healthy in space.

"Machines that work fine on Earth commonly break down in space"

NASA also has a new 'bioastronautics roadmap', released in February, that outlines 45 risks to space travellers that need further study. John Charles, deputy chief scientist for bioastronautics at the Johnson Space Center near Houston, Texas, calls it "our best collective guess" on the hazards of going to Mars, from kidney stones to contaminated water.

Some risks, such as the space sickness that afflicts two-thirds of shuttle astronauts, don't worry Rhea Seddon, assistant chief medical officer at the Vanderbilt Medical Group in Nashville, Tennessee, and a veteran of three shuttle missions. "Do you research it to death?" she asks, or simply administer the anti-nausea drug Phenergan, which works — although no one knows how.

Other risks are more serious. Four might scupper a Mars expedition: radiation, bone deterioration, psychosocial problems, and how to provide medical care in a weightless spacecraft millions of miles from home.

At least three can be researched partly or wholly on Earth. Psychosocial questions can be studied in analogous settings such as underwater habitats. NASA's radiobiology research has shifted to studies of ionizing radiation on animals at the Brookhaven National

spaceflight (T. Lang *et al.* *J. Bone Miner. Res.* **19**, 1006–1012; 2004). And even if some astronauts gain overall bone mass, loss of some kinds of bone tissue might weaken their limbs. Anecdotes aside, says Pawelczyk, bone loss remains a serious concern.

Seddon agrees, and says this is one area where more research really is needed. "Ed Lu is probably willing to go to Mars tomorrow," she says. "I wouldn't be."

The station is a good place for such research, but there is one major problem: not enough test subjects. Today the station can accommodate just three long-term residents — the number that fit in the Russian Soyuz craft that doubles as a lifeboat. When the shuttle was grounded following the Columbia disaster, NASA cut the crew to two to save on resources.

According to NASA's bioastronautics roadmap, addressing all 45 risks would require flying 200 test subjects on the station. Charles calls that "an embarrassing number, because it's way beyond the realm of possibility". So far there have been 26 long-term residents in five years.

Attaching two Soyuzes to the station would raise the crew to six, but there is currently no timetable, nor funding, for doing so. And NASA's own replacement vehicle for the shuttle wouldn't be ready to operate as a six-person lifeboat until at least 2010.

One solution is to supplement human data with animal studies. Earlier plans called for a large centrifuge, built by Japan, as part of a programme of gravitational biology research. But NASA is reportedly preparing to cut animal research from the station. Funding for the centrifuge and animal habitats will therefore be one of the most closely scrutinized items in the space-station research plan unveiled later this month.

Seddon understands NASA's reluctance to send animals into space. They are expensive and difficult to house in weightlessness, and they complicate astronaut training. But they may be the only way to address one of NASA's top concerns about going to Mars.

Lu, meanwhile, highlights a different worry. He wants to make absolutely sure the equipment works — particularly the life-support system that cleans the astronauts' air and filters their water. He knows from personal experience how commonly machines that work fine on Earth break down in space. If the air filtration on a Mars ship stops working, he says, "you're dead about two months out".

That's the real problem for long-term human spaceflight, thinks Lu, more than stuffy noses or weak bones. Testing technology may end up as the station's final purpose. "The actual ship itself is the experiment," he says. "I think that's really what the station has to offer."

Tony Reichardt writes for *Nature* from Washington DC.

Additional reporting by Jenny Hogan and David Cyranoski.

Stalled countdown



ESA/ALenia

Hitching a ride: Europe's Columbus lab needs a lift to the space station on the shuttle.

While NASA decides what research to do on the space station, the European and Japanese space agencies are still waiting for their laboratories to launch.

Europe's Columbus module, currently stored in Bremen, Germany, is ready to go. The Japanese are putting the finishing touches to their Kibo laboratory. Both need to hitch a ride on the beleaguered shuttle.

Columbus and Kibo are equipped for basic science — including studying plants and animals in microgravity, and for investigating material growth and fluid physics. Neither agency has been swayed by NASA's overhaul of the science goals to focus on exploration.

"If anything, we might try to accommodate some of the US scientists on our programme," says Marc Heppener, who coordinates station research for the European Space Agency (ESA). "We see, with regret, that some of their good science proposals are not making it to the space station."

Europe's research plans are in good health, says Heppener, despite frustrating delays to the shuttle schedule. Some European experiments have been rescheduled to fly on Russian Soyuz rockets, and the agency has funded a vigorous ground-based research programme.

So any further hold-ups generated by Discovery's recent troubled flight won't derail the European effort. "We've waited so long, we can wait another few months," says Alan Thirkettle, of ESA's Directorate of Human Spaceflight.

But Europe is considering how best to modify its research if the partially constructed space station is stuck with two or three astronauts, rather than the hoped-for crew of six. "Plan B is not to start panicking," says Thirkettle. Instead, experiments that require a lot of time will be scrapped and others automated.

Japan is also sticking to plan A. "If the United States won't launch Kibo, which is a possibility, we don't have any alternative to go to," says Yasunori Matogawa, an associate executive director at the Japanese agency.

NASA is expected to present options to its partners by the end of the month. **J.H. & D.C.**



NASA

Short-staffed: the International Space Station is currently home to just two astronauts.

Laboratory in New York. And in-space medical care could be practised on the ground, with notable exceptions such as weightless surgery.

Attempts to curb bone loss could benefit from research on the station — but how big a risk is it really? Based on recent astronauts' experience, Lu thinks diligent exercise will hold the problem at bay. Combining workouts with bisphosphonates, a class of drugs developed to treat osteoporosis, could be the answer.

Not so fast, says James Pawelczyk, a Pennsylvania State University neurophysiologist who flew on a shuttle Spacelab mission in 1998. Research published by Thomas Lang of the University of California, San Francisco, based on data from 14 space-station astronauts, shows substantial loss of bone in the hip and a smaller loss in the spine as a result of long-term

Order out of chaos

Can the behaviour of complex systems from cells to planetary climates be explained by the idea that they're driven to produce the maximum amount of disorder? **John Whitfield** investigates.

In the mid-1970s, the love affair between climatologists and computer models was beginning to blossom. By breaking down the atmosphere and ocean into ever-smaller interacting chunks in simulations, researchers found that they could mimic the behaviour of the global climate with reasonable success. But for Garth Paltridge, a climate scientist at the University of Tasmania in Hobart, Australia, these general circulation models (GCMs) were coming at the problem from the wrong direction. "I felt it was like trying to describe the behaviour of a gas by following the path of every molecule," he says.

Instead, Paltridge decided to look for a simple, general principle that might explain the climate as a whole — similar to the physical laws that predict the behaviour of a gas as the average state of its countless molecules. He focused on the concept of entropy, a measure of the disorder in a system created as it does work. His idea that the climate maximizes its entropy production¹ stirred a flurry of interest at the time. But it did little to halt the GCM bandwagon. And in time, even Paltridge stopped working on it.

Now he is back on the case. In the past couple of years, Paltridge's hypothesis of maximum entropy production (MEP) has been given a new theoretical underpinning. And although it's early days, researchers are exploring the concept as an explanation of the behaviour of complex systems, from the climate to cells, organisms, ecosystems and economies². Entropy could even explain how linked complex systems interact, which could potentially lend legitimacy to the contentious theory of Gaia — the idea that living things act together to regulate Earth's climate to keep conditions favourable for life.

Cycle of violence

Paltridge's original model was very simple: splitting Earth into ten regions, it used only a few parameters, such as the strength of solar energy and Earth's reflectivity. More solar energy falls on the equatorial region than on the poles, and our weather results from the redistribution of this energy, through winds, currents and water vapour. Paltridge found that if he maximized the rate at which the atmosphere and oceans dissipated energy, his model world generated temperatures and cloud cover very similar to those seen for real. Our climate, he argued, creates weather that is as violent as possible, given the amount of energy available.



J. SCHWALTZ/MODIS/NASA/GSFC

Whipping up a storm: our climate opts for the most violent weather it can, given the energy available.

The problem was that there was no clear theoretical justification for why this should be so. The second law of thermodynamics states that a closed system will arrive at a state of maximum entropy, but it says nothing about how quickly it will get there, or about how much entropy a system such as the climate, which experiences a constant and massive input of energy, will produce. Over the

years, Paltridge and a few others tried to find a theory for MEP, and failed. Perhaps the resemblance between our climate and a system tuned to generate maximum disorder was just chance.

But the discovery that MEP could apply to atmospheres besides Earth's made it seem less of a coincidence. In 2001, Ralph Lorenz, a planetary scientist at the University of Arizona

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Snow hope: Japanese researchers are using a model avalanche of styrene beads (inset) to see if entropy theory can describe the real thing.

in Tucson, was trying to model the climate on Saturn's moon Titan. Titan is smaller than Earth, has a thicker atmosphere, and turns more slowly. This ought to mean that heat moves quickly through its atmosphere, making polar and equatorial climates very similar. Yet Titan's equator is 4 °C warmer than its poles. Only when Lorenz applied an MEP model could he produce the right temperature difference³. The same approach also gave accurate predictions of the winds and carbon dioxide frosts on Mars, which has a much thinner atmosphere than Earth and whose weather had similarly stumped conventional climate models.

MEP is not the same as maximizing the rate of heat transport. A system's entropy production depends on its energy input, and — crucially — on the temperature difference between its interacting parts. Very rapid heat transport would level out the temperature gradient between Titan's equator and poles, and so reduce the amount of work that its atmosphere could do, as if a ball were rolling slowly down a very gentle slope. Very slow heat transport would create a large temperature gradient, but the system would be close to equilibrium, and little work would be done — like a ball trying to roll down a steep but sticky slope. MEP needs a climate somewhere between these two extremes. Lorenz became



an entropy enthusiast. "I was evangelical about it — it made so much intrinsic sense," he says. But still the theoretical problem remained.

The breakthrough came in 2003, when Rod-eck Dewar, a theoretical physicist turned ecosystem modeller working for INRA, the French agricultural research agency, in Bordeaux, turned to information theory — a branch of mathematics dealing with communication and uncertainty. Thermodynamics can be expressed in terms of information theory: in the 1950s, it had been shown that the entropy of a system in equilibrium, such as a sealed container of gas, can be reformulated as a measure of missing information.

Out of order

Dewar extended the theory to non-equilibrium systems — such as the planetary climates modelled by Paltridge and Lorenz. In essence, he showed that what is true for a small container of gas molecules, with no energy input, should also apply if you put an astronomical number of gas and water molecules into an atmosphere-sized container and heat them up.

Dewar's theory says that for a large, complex system, the state of MEP is the most probable sum of its microscopic parts^{4,5}. One need know only the constraints that influence the behaviour of the whole system, and nothing about the seething complexity of all its constituent parts because, at the large scale, all the different microscopic arrangements look the same. "What we see at the macroscopic scale is the most probable behaviour, because it can be realized in the greatest number of ways microscopically," Dewar says.

But the theory comes with strings attached. First, the system must also be free to 'choose' between different states. For the climate this should be no problem: it has myriad possible configurations, from the local changeability of wind and clouds, to switches between ice ages and interglacial periods. But the theory also applies only to systems in a steady state: that is, enough energy must be passing through to preserve large-scale structure, but not so much, or so little, that this structure is disrupted. For the climate, this might not always be the case: a world that was rotating very rapidly; had a very thin atmosphere; which was in a glaciated, 'snowball' state; or which was experiencing strong man-made climate change, might be prevented from settling into MEP.

These constraints may limit the utility of MEP for climate modelling. Although the climate might be in a steady state on a planetary scale, and over long time-frames, at other scales things are always changing. What's more, ice, clouds and the ocean at different depths and

T. G. LAMAN/NATL GEOGRAPHIC/GETTY IMAGES; Y. NOHGUCHI/H. OZAWA (INSET)

latitudes respond to the same forces at vastly different rates. "It's not just the end result that's important, it's how you get there," says Kevin Trenberth of the National Center for Atmospheric Research in Boulder, Colorado.

Lorenz agrees that MEP is unlikely to revolutionize the way we model the climate — but he believes the idea could also be used to gauge the trustworthiness of GCMs. If a model's output is very far from MEP, it should throw up a warning flag, he suggests. Where MEP theory will be most useful, he adds, is as a means of getting a broad-brush picture of climates about which we know very little, such as those of extrasolar planets, or in Earth's distant history. MEP models allow surface temperatures across a planet to be calculated from knowing only how much of its star's light falls on it, how much is reflected back into space, the tilt of the world's axis, and its absorbance of infrared radiation. All these parameters can in principle be measured with a telescope⁶.

Natural disarray

Now researchers are looking for other systems that might be amenable to modelling using Dewar's theory. Plate tectonics, in which rock dissipates heat generated by radioactivity as it flows, might be complex enough to attain MEP⁶; so might the growth of crystals⁷. And Hisashi Ozawa, a climatologist at Hiroshima University in Japan, who has previously studied the entropy production of ocean currents⁸, is now staging small avalanches of powder in his laboratory, to see if MEP can explain the patterns that form at the front of the sliding mass. "There are many natural phenomena to which the MEP hypothesis can be applied," Ozawa says. "It's attractive because it's very general: it doesn't depend on specific physical and chemical properties."

Lorenz wonders about applications in economics. Money and goods flow between people in a similar way to heat in the atmosphere, he says: "I see a direct analogy between temperature gradients and price gradients." But economists aren't getting carried away just yet. "The deliberate nature of human actions makes economic systems qualitatively different from the climate, and adds an extra layer of complexity," says economist Matthias Ruth, of the University of Maryland in College Park.

Dewar is now working to apply MEP theory to biological systems ranging from cells to the planet. At a small scale, he is trying to see whether the theory can explain when plants open and close their stomata, tiny pores in the leaves that regulate the flow of gases and water vapour in and out. This has traditionally been explained in terms of the plant's efforts to maximize its photosynthesis and minimize its water loss; perhaps it might be reinterpreted as the leaf maximizing the entropy produced by gas exchange. If so, this would indicate that the biologically optimal state is also, in physical terms, the most probable state.

At the ecosystem level, ecologists have long

used a metric called the Shannon diversity index, based on the number of species in a given place and their relative numbers. This index is mathematically identical to a measure of entropy. Dewar suspects that MEP theory might be able to predict the number of species in a place based on its energy input, and possibly explain why the places with most energy, the tropics, are also the most diverse — or, in other words, entropic.

At the global scale, Dewar and his colleagues are trying to put numbers on the fluxes of solar energy, heat, water vapour and carbon dioxide between plants and the atmosphere. Previously, these have been added to climate models as ad hoc fudge factors: MEP might allow them to be calculated from fundamental principles.

Total anarchy

By linking vegetation and the climate, MEP also offers a new twist on the Gaia hypothesis. The problem for Gaia theorists has been explaining why Earth's organisms should en masse 'want' to maintain a stable climate. MEP might offer a way out, says Axel Kleidon, a biogeophysicist at the University of Maryland. According to his models, provided a planet is suitable for life in the first place, biological activity increases the entropy production of the entire planetary system, both living and non-living. It also increases the number of states that the system can adopt, he says, so making MEP more likely⁹. This would give a version of Gaia in which life isn't manipulating the climate to its own ends. Instead, if both climate and ecosystems tend to a state of MEP,

the stability of the climate becomes a by-product of this state, towards which the system will return when perturbed.

Yet the concerns voiced regarding other applications of MEP theory still apply. Is life, at the planetary scale, in a steady state? And is it free to enter maximum entropy production? Based on his modelling, Tim Lenton of the University of East Anglia in Norwich, UK, one of the leading lights of Gaia research, is sceptical. "Gaia would benefit from a sound theoretical footing, but there's still a gap between Dewar's very elegant theoretical treatment and many of the systems we're interested in," he says.

Ecosystem modeller Marcel van Oijen of the Centre for Ecology and Hydrology in Edinburgh is also unsure whether MEP is compatible with what we know about evolution: natural selection might be a force strong enough to prevent life from attaining maximum entropy production. "The problem of reconciling MEP and natural selection will be a focus of debate in coming years," he says. "I see adaptation and constraints everywhere in biology, and these aren't accounted for in the derivation of MEP"

Dewar is undaunted, and predicts that a new order will arise from the current chaos, as theorists wrestle with the various potential applications of MEP. "The underlying theory suggests that it's entirely general," he says. "It's an organizational principle that potentially unifies biological and physical processes."

John Whitfield is a science writer in London.

"Maximum entropy production is an organizational principle that potentially unifies biological and physical processes."
— Roderick Dewar

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The opening and closing of a leaf's pores (pictured) may be governed by entropy.

1. Paltridge, G. W. Q. J. R. Meteorol. Soc. **101**, 475–484 (1975).
2. Kleidon, A. & Lorenz, R. D. (eds) *Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond* (Springer, Heidelberg, 2005).
3. Lorenz, R. D., Lunine, J. I., Withers, P. G. & McKay, C. P. *Geophys. Res. Lett.* **28**, 415–418 (2001).
4. Dewar, R. C. J. *Phys. A* **36**, 631–641 (2003).
5. Dewar, R. C. J. *Phys. A* **38**, L371–L381 (2005).
6. Lorenz, R. D. *Int. J. Astrobiol.* **1**, 3–13 (2002).
7. Hill, A. *Nature* **348**, 426–428 (1990).
8. Shimokawa, S. & Ozawa, H. Q. J. R. Meteorol. Soc. **128**, 2115–2128 (2002).
9. Kleidon, A. *Climatic Change* **66**, 271–319 (2004).



The life aquatic

Cindy Lee Van Dover likes nothing better than to be on the ocean floor. **Emma Marris** meets the unconventional biologist who has devoted her life to studying the exotic ecosystems of the deep.

Some 2,300 metres under the sea, Richard Lutz was beginning to wonder whether he should worry. The deep-sea submersible *Alvin* in which he was riding had come to a shuddering halt. Next to him, the pilot stared intently out of the windows.

"We can't move," said Cindy Lee Van Dover, at the controls of a truck-sized craft stuck at the bottom of the ocean. "We are just going to sit here and think."

As Lutz, a marine biologist at Rutgers University, tells it, Van Dover's calm approach got them out of a potentially deadly situation that day in 1991. He was readying his handheld tape recorder to capture his last words, but after checking and rechecking *Alvin's* controls,

Van Dover concluded that the sub's front shield had probably scooped up hundreds of kilograms of mud. Hours later, she had managed to ditch enough weight to allow the sub to ascend safely to the surface.

Such deliberate cool in the face of danger is practically a requirement for deep-sea pilots. But it has served Van Dover equally well in her unconventional academic career studying deep-sea ecosystems. In a small field crowded with powerful personalities, she has sometimes irritated the establishment by not following a traditional path. Yet her trademark intensity — a quality she readily describes as "stubbornness" — has often helped her to ride out the storms.

Van Dover has racked up hundreds of hours in submersibles on the sea floor, including a period in the early 1990s when she spent more time on the ocean bottom than any other scientist. That experience, combined with her natural curiosity for all things oceanic, set the stage for her to make leaps of reasoning that might seem counterintuitive to her colleagues, but which sometimes result in startling new insights into vent creatures.

These creatures thrive in a peculiar environment that, at first glance, seems hostile to life. They live along cracks in the sea floor where water, superheated by volcanic activity, gushes out of towering chimneys. These hydrothermal vents are home to communities of fantastic organisms ranging from mats of bacteria living off sulphides to long, pale, red-tipped tubeworms that sway in the current.

All at sea

An invertebrate zoologist at heart, Van Dover finds this diversity inspiring. She is excited by, as she puts it, "all the weird ways you can be alive". Her explorations have taken her from the relatively well characterized vents on the mid-Atlantic ridge, to the rarely visited depths of the Indian Ocean and the seas around Fiji and Easter Island.

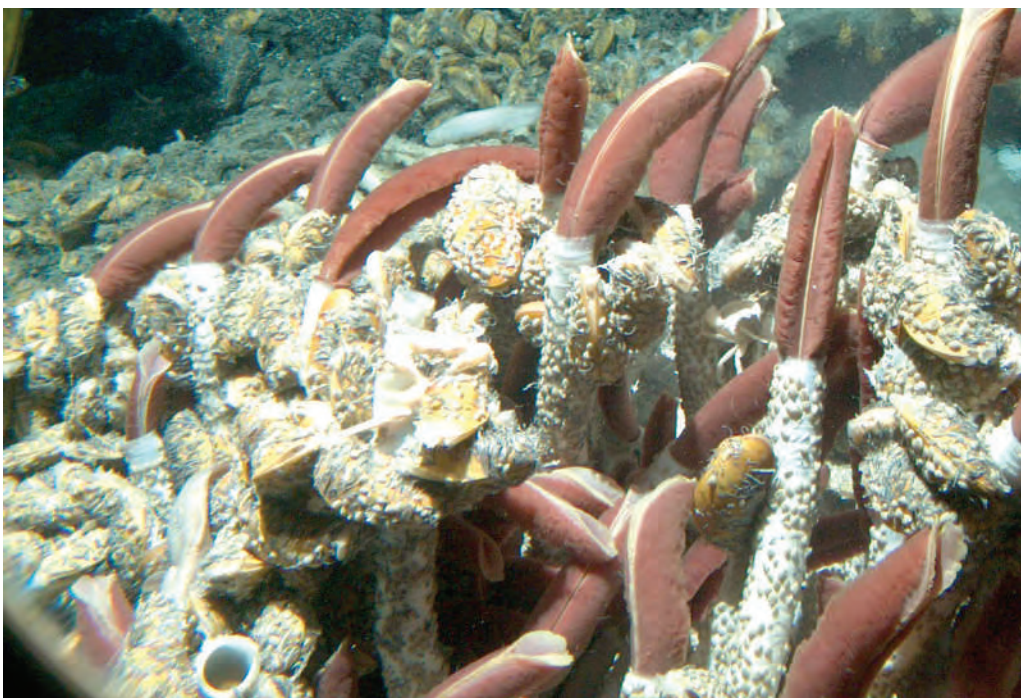
Yet despite all the time she has spent on — and in — the water, Van Dover still gets seasick on the first few days of a voyage. Not that she lets that put her off. "I love the romance of going to sea," she says with a smile. "I love the phrase, 'I'm going to sea'."

This life-long passion arose when she was an undergraduate studying zoology at Rutgers. Fascinated by the bizarre biology of deep-sea creatures, she applied to the joint graduate programme in oceanography run by the Massachusetts Institute of Technology and the Woods Hole Oceanographic Institution (WHOI). When she didn't get in, she spent six years working as an itinerant lab technician and sharpening her mathematical skills until she was accepted on to the programme.

At the WHOI, Van Dover was soon making waves with a seemingly outlandish hypothesis: she believed that there may be a light source in the inky depths by the vents. Van Dover had been puzzling over a species of eyeless rift shrimp called *Rimicaris exoculata*, which had been recovered from the Atlantic. Watching videos of the shrimps at the vents, she saw two reflective patches running down their backs. But when she examined the specimens brought to the surface, she found that the preservation process had rendered the blotches invisible.

The patches, which had never been noticed before, looked like some kind of sensor. Van Dover guessed that they might be highly modified eyes for sensing light. But why would a shrimp in the lightless depths need eyes?

To find out if there was any light to respond to, Van Dover asked researchers going on a vent-mapping expedition to use a sensitive camera, designed for astronomical imaging, to take pic-



Regular trips piloting *Alvin* (right) have allowed Cindy Lee Van Dover (left) to study the rich ecosystems found at deep-sea vents.

tures of the seemingly pitch-black chimneys. Sure enough, the passengers aboard *Alvin* sent a brief message to the surface: "Vents glow."

Van Dover suggested that the shrimps use their 'eyes' to keep them close enough to the vents to feed on bacteria, but far enough away to prevent them from being cooked by the ferocious heat¹. Other biologists gradually accepted the idea, convinced by her marshalling of the evidence. "I learned a lot from seeing those eyes and learning how to prove that they are eyes," she says. "You can't do science by assertion."

Driving force

The shrimp discovery propelled Van Dover into the media spotlight, although some within the small community of vent biologists were less than thrilled by the amount of attention she attracted. "I had a mentor who was not happy," she says. Not for the last time, she found herself at odds with some senior researchers in her field. And her next move took her colleagues even more by surprise.

In 1989, with her PhD fresh in her hand, Van Dover decided that the best way to get to the sea floor regularly would be to become a pilot for *Alvin*. Owned by the US Navy and run by the WHOI, the submersible has made more than 4,000 dives since it was commissioned in 1964, including a visit to the wreck of the *Titanic*.

To learn the sub's inner workings, she began memorizing blueprints and perfecting electrical repairs. It was not an easy job. Some of *Alvin*'s other pilots did not welcome her into their exclusively male club. Many of her academic colleagues thought she was throwing away a promising career in research. But Van Dover persevered, determined to spend as much time

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

as she could at the bottom of the sea.

After being certified as *Alvin*'s first and so far only female pilot, Van Dover spent three years nearly constantly at sea. She dived as often as every other day, spending her hours out of the water monitoring other dives from the boat or repairing the sub. Although the hours were long and the work tough, she says she got an unparalleled education: "I was essentially a postdoc to everybody who came aboard that ship."

Thanks in part to her pilot experience, Van Dover has visited the ocean floor more than 100 times, 48 as pilot-in-command. Although she never gets bored of the destination, the lengthy descents have become tiresome. "I wish I could just snap my fingers and be at the bottom," she says. These days, she often gives up her seat in the sub to a student.

Van Dover retired as a pilot in 1991, and now serves on a committee overseeing the building of a new sub to replace *Alvin*. But she remains fond of the vehicle she once knew so intimately. "She's a sweetheart," Van Dover says. "I still go kick her tyres when I'm anywhere near her."

Conquering *Alvin* took an intensity of purpose that characterizes Van Dover's career. "She is just so focused," says Robert Vrijenhoek, a deep-sea biologist at the Monterey Bay Aquarium Research Institute in Moss Landing, California, who has sailed with Van Dover

many times. "She sets her mind on doing something and you collaborate with her, and by George it gets done."

After her stint as *Alvin* pilot, Van Dover wrote a popular-science book about the deep sea called *The Octopus's Garden*. She found that writing had a therapeutic effect, allowing her to work through the stresses of pilot training and of trying to scramble back onto the academic track. "It was catharsis," she says.

Making a splash

Many of her peers see her first as an ambassador of vent science. "A very eloquent writer who has done some great communicating to the general public," is how Charles Fisher, a vent biologist at Pennsylvania State University, describes her. "Her research is easy to explain and sexy, and there is nothing wrong with that."

Van Dover returned to academia in 1992, first at the WHOI, and later at Duke University in Durham, North Carolina, and the University of Alaska in Fairbanks. Now she is based at the College of William and Mary in Williamsburg, Virginia. When she isn't teaching, she is back at sea taking long cruises to the vents.

She admits that she works on whatever interests her most, and she rounds up whomever she needs to get the job done. "I've really been an opportunist," she says. "When I see a good story, I go for it."

Her most recent finding was published just this summer, and in some ways it picks up where she left off as a graduate student. While hunting for microscopic life around vents at the East Pacific Rise, Van Dover and her team found photosynthetic bacteria living in the black depths². If further research shows that the creatures live by the vents all the time, and don't just visit from brighter areas, then these bacteria would be the first organisms ever found to photosynthesize without sunlight. Astrobiologists are thrilled about this finding, envisioning distant dark planets teeming with photosynthetic life.

As for the source of the light at the vents, that remains a mystery. It can't be explained by thermal radiation from the vents alone, as too much of the light is in the visible part of the spectrum. Possible sources include sonoluminescence, in which imploding bubbles emit a brief flash of light, or sulphide oxidation by bacteria.

For now, Van Dover says she will leave that question for other specialists. She has moved on to study shellfish diseases — in particular a virulent condition that turns the flesh of giant vent mussels black. The myriad life-forms by the vents are likely to provide her career with plenty more twists and turns, but whichever path she takes it will lead only one way in the end: straight to the bottom.

Emma Marris is a Washington correspondent for Nature.

1. Van Dover, C. L., Szuts, E. Z., Chamberlain, S. C. & Cann, J. R. *Nature* **337**, 458–460 (1989).
2. Beatty, J. T. et al. *Proc. Natl Acad. Sci. USA* **102**, 9306–9310 (2005).

VAN DOVER LAB

R. WHITE/CORBIS

BUSINESS

Drug firms back-pedal on direct advertising

The pharmaceutical industry is taking a long, hard look at how it promotes its products to the public. Colin Macilwain reports.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The woman on the television has an enticing question: "In the mood for something different?" Levitra, she purrs, "is the best way to experience that difference".

Until this spring, her proposition was at the centre of one of the largest TV ad campaigns in America. In 2004, Bayer, Schering-Plough and GlaxoSmithKline spent \$157 million on advertising Levitra — a treatment for erectile dysfunction — directly to consumers. That's a pretty big investment for a product whose sales in the year amounted to about \$250 million.

But back in April, the US Food and Drug Administration (FDA) decided the seductive selling had to stop. It was, as the regulator unsexily put it, "in violation of the Federal Food, Drug and Cosmetic Act, 21 U.S.C. § 352(n), and FDA implementing regulations, 21 CFR §§ 202.1(e)(1) and (e)(3)". That means it didn't say what the drug was for, or mention its side effects. Most importantly, the regulator noted, there was no evidence that Levitra provides "better" performance than its rivals.

The FDA action was part of a wider backlash against some of the perceived excesses of direct-to-consumer advertising of prescrip-

Hard sell: advertising for the impotence drug Levitra was aimed directly at consumers.

tion drugs. Earlier this month, the industry faced up to the challenge by issuing a set of voluntary guidelines that it hopes will meet public concern about the ads and head off regulation by the government.

Under the guidelines, members of the industry association PhRMA will pause before advertising newly approved drugs to the public. "The centrepiece of this is that companies will spend time educating healthcare professionals before beginning an ad campaign," says Billy Tauzin, president of PhRMA.

The guidelines would have ruled out the Levitra ad on several counts. It was a 'reminder' ad — a short commercial that skips details of the illness to be treated, or of the drug's side effects — and the guidelines will bar these. It ran in the early evening when families were watching, but the PhRMA rules pledge to "avoid audiences that are not age-appropriate". Also, the code promises, all ads will "respect the seriousness of the health condition and the medicine being advertised".

"We think it is critical that ads create the right kind of impression," explains Peter Dolan, chief executive of Bristol-Myers Squibb. People in the pharmaceutical industry have "a higher calling" that its commercials should reflect, says William Weldon, chairman of Johnson & Johnson and of PhRMA's board. "Direct-to-consumer advertising has drawn fire from some quarters," he says, "and some of it is well-deserved."

Three factors drove the appearance of the guidelines, industry observers say. First of all, powerful members of Congress have been threatening regulation to control the adverts. Second, the public prestige of the drug industry has been declining, and some of that decline has been attributed to the advertising. And third, industry observers say, companies are worried that some commercials may open them up to litigation.

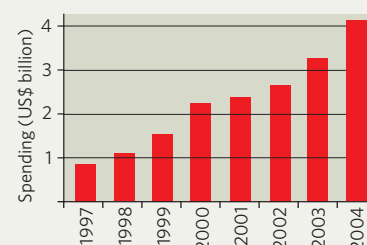
Bigger than Budweiser

The issue was brought to the fore by the withdrawal of the painkiller Vioxx by Merck last autumn, after it was linked to heart attacks and strokes. In 2000, according to a survey in *MedAdNews*, Vioxx was the drug with the largest direct-to-consumer advertising campaign in the United States, costing \$160 million, more than Budweiser (\$146 million) or Pepsi (\$125 million). Thousands of patients' representatives are now suing Merck over Vioxx (see *Nature* 436, 459; 2005).

Direct-to-consumer advertising for prescription drugs first appeared on television in the United States in 1997, when the FDA lifted a regulation that required more supplementary information than can be squeezed into a television commercial. The money spent on these ads has been rising quickly ever since (see graph, below). The United States and New Zealand are the only countries in the world that allow such ads — and New Zealand is widely expected to ban them later this year.

Companies argue that direct advertising helps patients get the therapies they need when they visit the doctor. "Advertising improves markets," says Jack Calfee, an economist at the American Enterprise Institute, who

DIRECT-TO-CONSUMER ADVERTISING FOR PRESCRIPTION DRUGS



spoke at a National Academies seminar on drug advertising on 3 August. "There's no reason to believe that pharmaceuticals are an exception to that," Calfee says that advertising provides information that consumers would not otherwise have, and helps them get the prescriptions they need from their doctors.

But the data are open to different interpretations. Richard Kravitz of the University of California, Davis, investigated how physicians respond to patients who walked into their surgeries to complain about depression (R. L. Kravitz *et al.* *J. Am. Med. Assoc.* **293**, 1995–2002; 2005). Like the drug ads, Kravitz's study provided employment for actresses, who feigned the symptoms of major depression or of a milder version known as adjustment disorder. On different visits, they asked for Paxil, an antidepressant made by GlaxoSmithKline; or for an antidepressant without naming one; or they didn't specifically ask for medicine at all.

"It was a complicated study, with some complicated findings," Kravitz says. Patients with the symptoms of major depression were much more likely to get the prescription that most doctors think they should get if they ask for a drug by name. But patients who reported the symptoms of adjustment disorder and don't need an antidepressant were also more likely to be prescribed one if they named Paxil.

Political pressure

Matt Hollon, a physician at the University of Washington who wrote an editorial accompanying Kravitz's finding (*J. Am. Med. Assoc.* **293**, 2030–2033; 2005), says it is not known whether the advertising influences people who really need treatment or gets the attention of those who don't. "That's the big, unresolved question about direct-to-consumer advertising," he says.

The reception for the industry's voluntary guidelines has also been mixed. "It's a welcome step but it is insufficient," says perennial industry critic Sidney Wolfe of the US consumer group Public Citizen.

Perhaps more significantly from the industry's point of view, powerful Republican politicians say the guidelines don't go far enough. Chuck Grassley (Republican, Iowa), chair of the Senate finance committee, says he will continue to back legislation requiring the ads to be vetted by the FDA before transmission.

Bill Frist (Republican, Tennessee) called the guidelines "an important first step" but repeated his call for a moratorium on advertising newly approved drugs.

But Tauzin says the guidelines will lead to ads that are "more educational, more balanced and more complete". If he's right, the public will experience a difference, and regulatory pressure on advertising could wane. ■

IN BRIEF

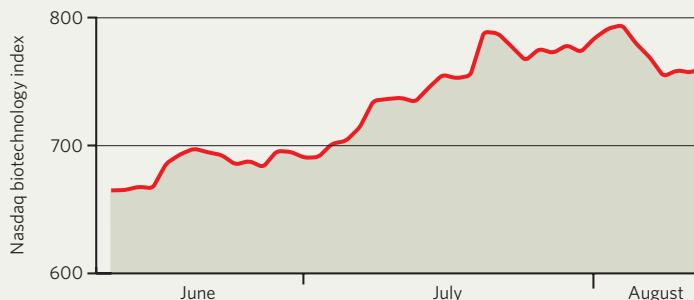
WALL STREET PROBE The Securities and Exchange Commission (SEC) has been asked to investigate claims that stock-market analysts are paying doctors to leak early findings from clinical trials of drugs. Senator Chuck Grassley (Republican, Iowa), chair of the Senate finance committee, has asked the SEC to look into the allegations, which were first reported on 7 August in *The Seattle Times*. The report said that analysts used the information to promote biotechnology stocks, whose values rest heavily on single trial outcomes.

OUT OF ORBIT The man in charge of satellite and intelligence programmes at aerospace company Boeing is to retire early, amid allegations of cost overruns on vast contracts to build secret networks of satellites for the US government. Roger Roberts, the senior executive responsible for the programmes, will retire at the age of 58. Information about one of the contracts — a \$20-billion project to build spy satellites for the National Reconnaissance Office — has leaked into the public domain as members of Congress grow concerned at spiralling costs.

GREATER ELAN Irish biotechnology company Elan and its US partner Biogen Idec have said that a safety evaluation of their multiple-sclerosis drug Tysabri is proceeding well and could be completed this summer. Shares in Elan rose by about a fifth to \$7.25 on the news. They had crashed earlier this year after it became known that two patients who had been taking part in the trials had died of a rare brain disease. Elan and Biogen are expected to resume large-scale trials of the drug soon.

MARKET WATCH

BIOTECHNOLOGY STOCKS



This week Wood Mackenzie, an Edinburgh-based research and consulting firm, reviews recent trends in biotechnology stocks.

Having been fairly flat since April, the Nasdaq Biotechnology Index twitched upwards in mid-June and then gathered momentum to show a sustained rise during July. Thanks to stronger than expected second-quarter financial results from several companies, the index rose by an encouraging 14% over the past two months.

In July, industry bellwether Amgen of Thousand Oaks, California, regained its crown as the largest single company in the sector by market capitalization, after announcing strong sales growth of its existing drugs, as well as positive clinical-trial results for several new ones. The tussle for the top market valuation between Amgen and Genentech of South San Francisco, California, has helped to bring investor attention to the sector as a whole, with

a positive effect on stock valuations.

Other companies reporting strong second-quarter results include Genzyme of Cambridge, Massachusetts, whose improved gross margins on drug sales led to increased earnings, and Kos Pharmaceuticals of Cranbury, New Jersey, which saw increased sales of its cholesterol and asthma drugs.

Protein Design Labs of Fremont, California, announced that it expects to achieve sustainable positive cash flow by the end of 2005. The company was also buoyed by the announcement of a strategic alliance with Biogen Idec of Cambridge, Massachusetts, to co-develop monoclonal-antibody drugs.

Despite the index's recent good performance, it has still merely regained the value it had at the beginning of the year. The industry needs to maintain its momentum for a few months more if it is to achieve a turnaround in its fortunes. ■

www.woodmac.com

Buddhism is no bar to an open mind. Is science?

SIR — I was appalled to learn, through your News story “Neuroscientists see red over Dalai Lama” (*Nature* **436**, 453; 2005), of a petition by a group of neuroscientists to cancel a lecture by the Dalai Lama scheduled for the November meeting of the Society for Neuroscience.

What are the motives behind this petition? If the research itself is controversial, then the controversies should be aired at this meeting. The Dalai Lama is fully capable of scientific discussion, as reported in an earlier News Feature “Buddhism on the brain” (*Nature* **432**, 670; 2005). Asked what would happen if neuroscience came up with information that directly contradicted Buddhist philosophy, the Dalai Lama is quoted as answering: “Then we would have to change the philosophy to match the science”. What is the harm in listening to a man, religious figure or not, who is so open-minded?

Robert Desimone’s statement that the Society for Neuroscience should “distance itself as much as it can from the Dalai Lama and his beliefs” is also unjustifiable. In the past year, the Dalai Lama has spoken in favour of scientific truth-seeking, human rights, species conservation and peace. If these are things that we, as scientists, need to distance ourselves from, then perhaps we deserve the damaging stereotypes we are given in the popular media.

Janis L. Dickinson

Hastings Natural History Reservation, Museum of Vertebrate Zoology, University of California, Berkeley, 38601 Carmel Valley Road, Carmel Valley, California 93924, USA

Power-plant design should prepare for carbon capture

SIR — Your News Feature “China’s burning ambition” (*Nature* **435**, 1152–1154; 2005) identifies gasification-based technologies as important for limiting future pollutant and CO₂ emissions from China, but it fails to consider another major requirement for any future large-scale reduction strategy. There was no mention of CO₂ capture and storage from the hundreds of new coal-combustion power plants that will inevitably be built in China during the coming decades. These modern power stations are likely to continue in use for about 50 years, and their combined emissions could exceed current UK emissions, for example, several times over.

As your News Feature notes, it would be very expensive to move immediately from combustion technologies to gasification-based options for all new coal plants. Fortunately, this is not necessary to prepare

for eventual CO₂ capture and geological carbon sequestration from these actual plants. Modern combustion plants using supercritical steam conditions, like those now being built in China, have coal-to-electricity efficiencies that at least match those of integrated gasifier combined cycle plants. With the latest technologies, efficiency penalties for capture are now also comparable.

Building a new supercritical steam plant to be ‘capture ready’, so that CO₂ capture can be added later with minimal cost and performance impacts, will add a negligible amount to the initial cost, but could reduce subsequent economic penalties for adding capture by perhaps a third.

With no market or regulatory drivers currently in China to make coal-combustion plants ‘capture ready’, the technology has received little or no attention there, although it is being actively considered by electricity utilities in Canada and the United Kingdom.

Jon Gibbins

Energy Technology for Sustainable Development Group, Mechanical Engineering Department, Imperial College London, London SW7 2AZ, UK

Answering the critics of Japanese whale research

SIR — We wish to respond to a Commentary article by Nicholas J. Gales and colleagues, “Japan’s whaling plan under scrutiny” (*Nature* **435**, 883–884; 2005).

The title of the article inappropriately uses the term “whaling”. Research on whales is conducted according to provisions of the International Convention for the Regulation of Whaling (ICRW) and is fundamentally different from commercial whaling.

It is true that the government issuing research permits may determine the sample size, and there is no requirement to change this on the advice of other governments. But Japan has submitted annual research plans to the Scientific Committee (SC) of the International Whaling Commission (IWC) for review and has incorporated amendments following any constructive suggestions from the SC. Gales and colleagues present their negative criticism of Japanese research as if this was the opinion of the SC. However, many positive comments have been made and reported by the SC over the years.

The authors condemn Japanese research on the basis of the number of whales captured since 1987. In planning our research, we must carefully determine the sample size that can achieve statistically valid results while also safeguarding population levels. Any criticism should address the rationale for calculating sample size, not merely highlight a number that seems large. Our research proposal describes this process clearly.

Criticism that science is being used as a

cloak to hide the purpose of killing whales is inappropriate. The lethal method used for sampling is required to achieve our research objectives, including determination of age and detailed data on stomach contents. Also, parts of Japan’s research programme use non-lethal techniques for sighting surveys and oceanographic studies, as well as biopsy sampling.

Contrary to the authors’ comment that the publication record of our 18-year research programme is “very poor”, we have made more than 150 scientific papers available to the SC and had a further 79 published in academic peer-reviewed journals. Unfortunately, many journals reject papers that report data from lethal whale sampling, even though they accept papers on other lethally sampled mammals.

In January 2005, the government of Japan held a meeting to review data and results from the previous 17 years of research (JARPA) in order to assist the planning of the new research programme (JARPA II). Many anti-whaling scientists, including the authors of the Commentary article, chose not to attend, although every member of the SC was invited. The Commentary claims that Japanese whale research is being conducted in an area designated as a sanctuary by the IWC. But the creation of this sanctuary was not recommended by the SC, and it was designed only to protect whales from commercial whaling.

Our new plan for Antarctic whale research (JARPA II) has been designed with far-reaching objectives to ascertain the dynamics of the Antarctic ecosystem. The Commentary authors say this is a job for the Commission for the Conservation of Antarctic Marine Living Resources, but this organization does not conduct research on whales; Japan plans to collaborate with it to achieve JARPA II’s research objectives. Japan has provided skilled labour, logistics and vessels to the IWC for large-scale Antarctic whale surveys conducted with multinational researchers.

Research of the magnitude of JARPA II is costly. Funds may be obtained by selling the by-products of whale research, according to Article 8 of the ICRW. Critics who say this makes the research a commercial enterprise are denying a provision of the ICRW.

Finally, the Commentary was written and published using information from the SC that, under IWC rules, should have remained confidential until the IWC’s opening plenary session on 20 June.

Hiroshi Hatanaka

Institute of Cetacean Research, 4-5 Toyomi-Cho, Chuo-Ku, Tokyo, 104-0055, Japan

The Institute of Cetacean Research was offered an early opportunity to balance the Commentary published on 16 June with an accompanying article in the same issue, but declined *Nature’s* offer — Editor, *Nature*.

COMMENTARY

Re-wilding North America

A plan to restore animals that disappeared 13,000 years ago from Pleistocene North America offers an alternative conservation strategy for the twenty-first century, argue **Josh Donlan** and colleagues.

North America lost most of its large vertebrate species — its megafauna — some 13,000 years ago at the end of the Pleistocene. And now Africa's large mammals are dying, stranded on a continent where wars are waging over scarce resources. However much we would wish otherwise, humans will continue to cause extinctions, change ecosystems and alter the course of evolution. Here, we outline a bold plan for preserving some of our global megafaunal heritage — one that aims to restore some of the evolutionary and ecological potential that was lost 13,000 years ago, and which offers an alternative vision for twenty-first century conservation biology.

Our vision begins immediately, spans the coming century, and is justified on ecological, evolutionary, economic, aesthetic and ethical grounds. The idea is to actively promote the restoration of large wild vertebrates into North America in preference to the 'pests and weeds' (rats and dandelions) that will otherwise come to dominate the landscape. This 'Pleistocene re-wilding' would be achieved through a series of carefully managed ecosystem manipulations using closely related species as proxies for extinct large vertebrates, and would change the underlying premise of conservation biology from managing extinction to actively restoring natural processes.

Historic vision

Our proposal is based on several observations. First, Earth is nowhere pristine; our economics, politics, demographics and technology pervade every ecosystem. Such human influences are unprecedented and show alarming signs of worsening. Second, environmentalists are easily caricatured as purveyors of doom and gloom, to the detriment of conservation. Third, although human land-use patterns are dynamic and uncertain, in some areas, such as parts of the Great Plains in the United States, human populations are declining¹ — which may offer future conservation opportunities. Fourth, humans were probably at least partly responsible for the Late Pleistocene extinctions in North America, and our subsequent activities have curtailed the evolutionary potential of most remaining large vertebrates. We therefore bear an ethical responsibility to redress these problems.

North American conservationists routinely turn to the arrival of Columbus in 1492 as a

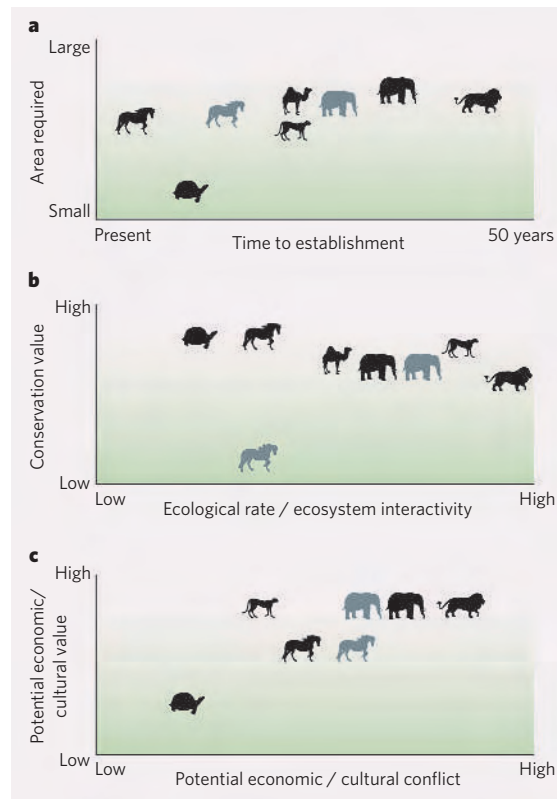


Figure 1 | Pleistocene re-wilding in North America. Symbols represent horses (*Equus caballus* and *E. asinus* in black; *E. przewalskii* and *E. hemionus* in grey), Bolson tortoises, camels, cheetahs, Asian (grey) and African (black) elephants, and lions. **a**, The likely timescale and area required to restore proxies for extinct large vertebrates. **b**, Conservation value and ecological role (interactivity with other species) on the landscape. **c**, Potential economic/cultural value versus potential conflict.

restoration benchmark. But the arrival of the first Americans from Eurasia roughly 13,000 years ago constitutes a less arbitrary baseline. Mammal body-size distributions were similar across all continents before the Late Pleistocene, but subsequent extinction of most large species drastically altered those distributions in favour of smaller species.

In the Americas, where large-vertebrate losses were greatest, the subsequent changes were undoubtedly ecologically and evolutionarily significant. Large carnivores and herbivores often play important roles in the maintenance of biodiversity, and thus many extinct mammals must have shaped the evolution of the species we know today². For example, the pronghorn (*Antilocapra americana*) evolved

over four million years in North American grasslands that changed abruptly in the Late Pleistocene; the now-extinct American cheetah (*Acinonyx trumani*), a key predator, almost certainly shaped the pronghorn's astonishing speed³.

Beasts of old

Although historical perspectives have influenced modern conservation planning, existing programmes do not adequately address the evolutionary potential and long-term processes involved in restoring large-mammal diversity. Africa and parts of Asia are now the only places where megafauna are relatively intact, and the loss of many of these species within this century seems likely. Given this risk of further extinction, re-wilding of North American sites carries global conservation implications.

Moreover, humans have emotional relationships with large vertebrates that reflect our own Pleistocene heritage. More than 1.5 million people annually visit San Diego's Wild Animal Park to catch a glimpse of large mammals — more than the number of visitors to most US National Parks. So an understanding of ecological and evolutionary history, inspired by visits to private or public reserves

containing free-roaming megafauna, could strengthen support for conservation. Pleistocene re-wilding would probably increase the appeal and economic value of both private and public reserves, as evidenced by the restoration of wolves to Yellowstone National Park⁴.

We foresee several phases to Pleistocene re-wilding, some of which are already under way. The 50-kg Bolson tortoise (*Gopherus flavo-marginatus*) was widely distributed across the Chihuahuan desert until the Late Pleistocene. Today it survives only in a small part of northern Mexico and is critically endangered. A number of appropriate sites exist for re-introduction, including Big Bend National Park, Texas. And repatriation of captive Bolson tortoises to a private ranch in New Mexico

is currently under study. Restoring North America's largest surviving temperate terrestrial reptile to its prehistoric range could bring ecological, evolutionary, economic and cultural benefits, with no apparent costs (Fig. 1).

Likewise, horses and camels originated in North America, and many species were present in the Late Pleistocene. Feral horses (*Equus caballus*) and asses (*E. asinus*), widely viewed as pests in the United States, are plausible proxies for extinct American species. Also, given that most of the surviving Eurasian and African species are now critically endangered, establishing Asian asses (*E. hemionus*) and Przewalski's horse (*E. przewalskii*) in North America might help prevent the extinction of these endangered species and would restore equid species to their evolutionary homeland.

Similarly, Bactrian camels (*Camelus bactrianus*) in North America could provide a modern proxy for *Camelops*, a late Pleistocene camelid. Wild Bactrian camels are on the verge of extinction, currently restricted to the Gobi desert. Domesticated or captive camels might benefit arid North American ecosystems by browsing on woody plants that today often dominate southwestern US landscapes. With proper management, camels could provide economic benefits as well⁵. The overall benefits and disadvantages of horses and camels as proxies will depend on local contexts, and possibly on the presence of appropriate predators.

Free to roam

The second, more controversial phase of Pleistocene re-wilding could also begin immediately, with the maintenance of small numbers of African cheetahs (*Acinonyx jubatus*), Asian (*Elephas maximus*) and African (*Loxodonta africana*) elephants, and lions (*Panthera leo*) on private property. Many of these animals are already in captivity in the United States, and the primary challenge will be to provide them with naturalistic settings, including large protected areas of appropriate habitat and, in the case of carnivores, live prey.

The African cheetah, a close relative of the American cheetah, has only a modest chance of persisting in the wild in the next century. Breeding programmes are not self-sustaining, but some of the 1,000 captive animals could be used in re-wilding. Free-roaming, managed cheetahs in the southwestern United States could save the fastest carnivore from extinction, restore what must have been strong interactions with pronghorn, and facilitate ecotourism as an economic alternative for ranchers (Fig. 1).

Managed elephant populations could similarly benefit ranchers through grassland maintenance and ecotourism (Fig. 1). Five species of proboscids (mammoth, mastadon and gomphotheres) once roamed North America in the Late Pleistocene; today many of the remain-

ing African and Asian elephants are in grave danger. Elephants inhibit woodland regeneration and promote grasslands, as Pleistocene proboscids probably once did. With appropriate resources, captive US stock and some of the 16,000 domesticated elephants in Asia could be introduced to North America, where they might suppress the woody plants that threaten western grasslands. Fencing, which can be effective in reducing human–elephant conflict in Africa, would be the main economic cost.

Lions, which play a pivotal ecological role in the Serengeti, represent the ultimate in Pleistocene re-wilding for North America. They are increasingly threatened, with populations in Asia and some parts of Africa critically endangered. Replacing the extinct American lion (*Panthera leo atrox*), although challenging, has clear aesthetic and economic benefits (Fig. 1).

Among the objections to Pleistocene re-wilding is that the proposed proxies are not genetically identical to the animals that formerly existed in North America. And our vision might strike some as 'playing God'. Existing lions and cheetahs are somewhat smaller than their extinct counterparts, for example, and *Camelus* is different from *Camelops*. 'Same' is relative, however, as illustrated by the highly successful reintroduction of peregrine falcons (*Falco peregrinus*) in North America. Captive-bred birds from seven subspecies on four continents were used, yet there were no differences among the birds in subsequent breeding success⁶, and the subspecies now serve as a collective proxy for the extinct midwestern peregrine falcon.

More challenging objections to Pleistocene re-wilding include the possibility of disease transmission, the fact that habitats have not remained static over millennia, and the likelihood of unexpected ecological and social consequences of reintroductions. These issues must be addressed by sound research, prescient management plans and unbiased public discourse on a case-by-case and locality-by-locality basis. Well-designed, hypothesis-driven experiments will be needed to assess the impacts of potential introductions before releases take place. Large tracts of private land probably hold the best immediate potential for such studies, with the fossil record and research providing guideposts and safeguards. For example, 77,000 large mammals (most of them Asian and African ungulates, but also cheetahs, camels and kangaroos) roam free on Texas ranches⁷, although their significance for conservation remains largely unevaluated.

The third stage in our vision for Pleistocene re-wilding would entail one or more 'ecological history parks', covering vast areas of economically depressed parts of the Great Plains. As is the case today in Africa, perimeter fencing would limit the movements of otherwise free-living ungulates, elephants and large carni-

vores, while surrounding towns would benefit economically from management and tourism-related jobs. A system of similar reserves across several continents offers the best hope for long-term survival of large mammals.

Meeting the challenge

In the coming century, by default or design, we will constrain the breadth and future evolutionary complexity of life on Earth. The default scenario will surely include ever more pest-and-weed dominated landscapes, the extinction of most, if not all, large vertebrates, and a continuing struggle to slow the loss of biodiversity.

Pleistocene re-wilding is an optimistic alternative. We ask of those who find the objections compelling, are you content with the negative slant of current conservation philosophy? Will you settle for an American wilderness emptier than it was just 100 centuries ago? Will you risk the extinction of the world's megafauna should economic, political and climate change prove catastrophic for those populations remaining in Asia and Africa? The obstacles are substantial and the risks are not trivial, but we can no longer accept a hands-off approach to wilderness preservation. Instead, we want to reinvigorate wild places, as widely and rapidly as is prudently possible. ■

Josh Donlan is in the Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, New York 14853, USA.

Co-authors are Harry W. Greene of Cornell University; Joel Berger who is at the Teton Field Office, Wildlife Conservation Society; Carl E. Bock and Jane H. Bock of the University of Colorado, Boulder; David A. Burney of Fordham University, New York; James A. Estes of the US Geological Survey, University of California, Santa Cruz; Dave Foreman of the Re-wilding Institute, Albuquerque, New Mexico; Paul S. Martin of the Department of Geosciences, University of Arizona, Tucson; Gary W. Roemer of the Department of Fishery and Wildlife Sciences, New Mexico State University, Las Cruces, New Mexico; Felisa A. Smith of the University of New Mexico, Albuquerque; and Michael E. Soulé who is based at Hotchkiss, Colorado.

1. Lonsdale, R. E. & Archer, J. C. *J. Geogr.* **97**, 108–122 (1998).
2. Janzen, D. H. & Martin, P. S. *Science* **215**, 19–27 (1982).
3. Byers, J. A. *American Pronghorn: Social Adaptations and the Ghosts of Predators Past* (Chicago Univ., Chicago, 1997).
4. Duffield, W. J. & Neher, C. J. *Economics of Wolf Recovery in Yellowstone National Park*. Transactions of the 61st North American Wildlife and Natural Resources Conference 285–292 (Wild Mgmt. Inst., Washington DC, 1996).
5. Phillips, A., Heucke, J., Dorgers, B. & O'Reilly, G. *Co-grazing Cattle and Camels. A report for the Rural Industries Research and Development Corporation* (Rural Industries Research and Development Corporation, Kingston, 2001).
6. Tordoff, H. B. & Redig, P. T. *Conserv. Biol.* **15**, 528–532 (2001).
7. Schmidly, D. J. *Texas Natural History: A Century of Change* (Texas Tech Univ. Press, Lubbock, 2002).

Acknowledgements: We thank the Environmental Leadership Program, Lichen Foundation, Turner Endangered Species Fund, New Mexico Agricultural Station, Ladder Ranch, C. Buell, S. Dobrott, T. Gorton, M. K. Phillips and J. C. Truett for support and encouragement.

BOOKS & ARTS

Live long and prosper

Science can boost your chance of reaching a healthy old age — but don't hold your breath for immortality.

Fantastic Voyage: Live Long Enough to Live Forever

by Ray Kurzweil & Terry Grossman
Rodale: 2005. 400 pp. £12.99, \$24.95

The Life Extension Revolution: The New Science of Growing Older Without Aging

Philip Lee Miller & The Life Extension Foundation (with Monica Reinagel)
Bantam: 2005. 416 pp. \$26

Tom Kirkwood

In 1939 the British parliament passed a little-known piece of legislation, the Cancer Act, which outlawed the publication of any advertisement containing “an offer to treat any person for cancer, or to prescribe any remedy therefor, or to give any advice in connection with the treatment thereof”. The act was passed to protect the public from potential profiteers at a time when not much was known about the underlying causes of cancer, and when effective treatments, other than radical surgery, were unavailable. The Cancer Act remains in force today, and several successful actions have been brought in recent years.

It is interesting that similar legislation appears never to have been considered necessary to protect the public against claims of treatment that might stave off the ageing process. Such ‘treatments’ have long been sold, with classic cases involving potentially dangerous medical procedures, such as the transplantation into humans of monkey testicles or the injection of fetal cells from sheep or goats. For the would-be confidence trickster, the life-extension scam is one of the oldest games in town. We must assume that no one thought it necessary to legislate against such claims because, despite the evidence to the contrary, it was believed that nobody would take them seriously.

How times have changed. Both *Fantastic Voyage* by Ray Kurzweil and Terry Grossman, and *The Life Extension Revolution* by Philip Lee Miller and the Life Extension Foundation, show that greater scientific understanding of the ageing process has led to a rapid growth in ‘anti-ageing’ medicine. The cosmetics industry and a wide range of other companies now market products, techniques and advice that are advertised as delaying the effects of ageing

and even helping us to live longer. A buzz is in the air that significant life extension is just a few years away. Indeed, the subtitle of *Fantastic Voyage* — “Live Long Enough to Live Forever” — seems designed to make you think that you had better buy it and act upon its advice now. Otherwise you might miss the bus that will carry future generations into the land of the endless tomorrow.

Peel away the gloss, however, and these two books turn out to be rather humdrum contributions to the growing genre of ‘how to’ manuals that aspire to tell us “how to benefit from cutting edge science and add years to your life” and “how to extend the prime of your life and rejuvenate your body, mind and spirit”. Both books do a fair job of summarizing the current state of knowledge about factors that can affect the ageing process and about what can sensibly be done to increase your chances of living into old age in good health.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Don't believe the hype: do you look like you were born yesterday?

The books overlap extensively, which on the whole points to a reassuring consensus. The downside is that there is little in either book that is scientifically exciting or new. Both review briefly what is known about the ageing process before zeroing in on the usual suspects: fats, sugars, obesity, inflammation, exercise, heart health, sex hormones, mental stimulus, food supplements, stress and so on.

Each of them also indulges in a bit of fairly pedestrian speculation about what the future might hold in terms of stem-cell therapies, nanobots, cryonics and the like. In many places, the evidence to support the efficacy of the authors' recommendations is thin, sometimes tenuously so. But on the whole they offer sound advice which, if followed, is likely not only to do you some good physically, but also to make you feel positive about confronting your personal ageing challenges.

The books themselves are basically sensible, so why is there such unnecessary hype on their covers? More generally, why is there a pervasive sense that advocates of life extension must make preposterous claims about imminent longevity gains if they are to gain public notice? In part, the answer lies in the practice of some on the fringes of scientific ageing research who have upped the ante by making wholly unsupported extrapolations from work in cells and in simple animal models such as yeast, nematode worms and fruitflies. In part, it also lies in the fact that ageing research is still relatively young. There is a media hunger, as yet unchecked by widespread general knowledge of what is or is not scientifically plausible, for fountain-of-youth stories that titillate the public. The BBC, for example, earnestly reported a few months ago the laughable claim that the first human who will live to 1,000 years is 60 already. If even Auntie, as the BBC is affectionately known, can succumb to such arrant nonsense, what hope is there for the more excitable sections of the media?

The saddest thing about the misleading and ultimately unhelpful nonsense uttered by the fantasists in the life-extension lobby is that research advances are genuinely changing the way we regard the ageing process. We have learnt that, far from being genetically determined, ageing is much more malleable than we used to think. This is opening up some exciting approaches to trying to improve health in later life. Much further work needs to be done to transform these beginnings into genuinely effective interventions, but the first steps have been taken. We know, for example, that, in model organisms, boosting some of the mechanisms for cellular maintenance and repair can indeed extend life-span. This does not mean

CHRISTIAN DANKIN/SPL

that the same techniques will necessarily work in humans, because we know from comparative studies that humans are already endowed, for good evolutionary reasons, with much better maintenance systems than shorter-lived species. By analogy, a design modification that boosts the performance of my own modest car will not necessarily make a Maserati go faster,

as the Maserati is engineered for peak performance already. But we can try.

Did the 1939 Cancer Act play much part in creating the relatively mature discussion and media reporting of advances in cancer research? Given that it was specific to Britain, I suspect that a deeper common sense prevailed. Let us hope that similar common sense

can be harnessed to take us forward more responsibly than at present into a world in which life-span and health-span are both likely to increase further. ■

Tom Kirkwood is at the Henry Wellcome Laboratory for Biogerontology Research, Institute for Ageing and Health, University of Newcastle, Newcastle upon Tyne NE4 6BE, UK.

Anticlimax

The Case of the Female Orgasm: Bias in the Science of Evolution

by Elisabeth A. Lloyd

Harvard University Press: 2005. 320 pp.
\$27.95, £18.95

Olivia P. Judson

For men, orgasm is an intimate part of reproduction: ejaculation doesn't usually happen without it. Presumably, male orgasms evolved because, in the past, males who experienced sexual pleasure were more likely to have sex, and so were more likely to sire children. But what about orgasm in women? Women can become pregnant without orgasm; indeed, some women bear lots of children without ever experiencing one. So how has the female orgasm evolved?

There are two basic possibilities. The female orgasm may have evolved under natural selection on females, which is to say that females who have the capacity to reach orgasm have historically had more surviving children than females who do not. Alternatively, it may have evolved as a by-product of natural selection on something else. A number of evolutionary biologists have hypothesized about the former, imagining various ways that orgasm might have enhanced female reproductive success. Elisabeth Lloyd, a philosopher of science at Indiana University, prefers the second.

In *The Case of the Female Orgasm*, Lloyd champions the notion — first advanced by Donald Symons in 1979 — that orgasm in women is an accidental consequence of the fact that the clitoris develops from tissue that in a male embryo will become the penis. This would mean that women have orgasms just because men do, not because it enhances their reproductive success. Lloyd also mounts a scathing attack on those who have speculated about how orgasm might have been subject to natural selection on females. She accuses them of failures of logic, shoddy data analysis, and a tendency to ignore data they don't like. She says they commit these sins because they are hostage to a variety of unexamined assumptions, the most egregious being 'adaptationism' — an (in her view) absurd and unjustified commitment to natural selection as an explanatory force in evolution.

Lloyd cites several facts to support her contention that female orgasm is a by-product.

First, there are no data showing that women who reach orgasm during sex have greater reproductive success than women who do not; moreover, orgasm is unnecessary for conception. Second, during the past 50 years, surveys of Western women have found that although a minority always reach orgasm during copulation, some never do, and everyone else does only sometimes. Third, most women find it easier to reach orgasm through manual stimulation than through stimulation from the penis. Finally, some other female primates, such as the stump-tailed macaque, the bonobo and the chimpanzee, can reach orgasm.

But none of these comes close to dealing a hammer blow to natural selection. Consider the fact that the clitoris develops from the same tissue as the penis. This tells us something about the origin of the clitoris, but little about why it is still here. Once something has arisen, it can still be subject to natural selection. It may be that the clitoris has been modified to help women achieve orgasm. Then again, it may not: we don't know.

Or consider the matter of orgasm and reproductive success. There are no data showing that orgasm enhances reproductive success; but nor are there data showing that it doesn't. What conclusion can we draw? None: absence of evidence is not evidence of absence.

Or consider the fact that not all women experience orgasm during sex. Lloyd equates

variation in phenotype with proof that natural selection has not acted. But this need not be so: we all have eyes, yet we cannot all see equally well. No one would argue that eyes have not evolved under natural selection on vision.

The sad fact is that, for now, all statements about the evolution of the female orgasm are conjectures in an empirical vacuum. To advance the debate, we need data.

The most obvious approach would be to ascertain whether there is (or was) a link between orgasm and reproductive success. Measuring the relationship between a given trait and reproductive success is difficult in any organism. It is obviously impossible to know whether orgasmic women have tended to have more children than anorgasmic women. The best we can do is try to infer.

The fact that orgasm is not necessary for conception rules out the obvious way that orgasm could enhance reproductive success — but it could have more subtle effects. For example, could orgasm during sex induce ovulation? In mammals such as ferrets and cats, ovulation is induced by stimulation from the male; might it be facultatively induced in humans? As far as I know, such an effect has not been reported for any primate, but then, as far as I know, no one has looked for it.

We also need to know far more about the nature of orgasm. Orgasm is the result of two phenomena: contractions in the pelvic region,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

and the perception of pleasurable sensations by the brain. Yet we have little understanding of how the two components relate to each other. Moreover, a distinction is often made between clitoral and vaginal orgasm. Whether these are physiologically different — let alone whether they evolved under different selection pressures — is unknown. Indeed, the neuroanatomy of the genital region is poorly understood, and we have scant data on how much it varies among women. Brain scans suggest that different parts of the brain may be involved in orgasm for males and females — which would be consistent with natural selection acting on females directly — but the sample sizes are as yet too small to draw confident conclusions.

To understand the significance of the varia-

tion in women's experience of orgasm, we need to know what causes this variation. Is it due to genetic differences in genital anatomy? To differences in the way brains perceive pleasure? To psychological or cultural factors? Or to a physical incompatibility between a woman and her partner(s)? In other words, do some women lack a capacity for orgasm, or is the capacity there but never realized? Again, data are lacking. A recent twin study (K. M. Dunn *et al. Biol. Lett.* doi:10.1098/rsbl.2005.0308; 2005) suggests there may be a genetic component — if your identical twin has orgasms it's likely that you do too — but whether the effect is due to physiology or psychology is unclear.

And we need to know more about when other primates experience orgasm. Do females

in other species have orgasms with some males but not with others? No one knows. Here, too, we need to know how pelvic contractions translate into brain waves. And we need to investigate males as well as females: it is often simply assumed that males in other species have orgasms. Data on other primates will help us to understand the relationship between male and female orgasm, and whether the female orgasm evolved before the split between humans, chimpanzees and bonobos.

In short, it's time to collect data. Without it, the debate will remain like sex sometimes is: furious, empty and anticlimactic. ■

Olivia P. Judson is a research fellow in evolutionary biology at Imperial College, Prince Consort Road, London SW7 2BW, UK.

Prussian precision

Anton Hallmann's technical drawings brought geometry to life.

Martin Kemp

Technical drawing lay at the heart of a flourishing of both the arts and the physical sciences in nineteenth-century Berlin. The precise lines of geometry obtained practical expression and gave shape to painting, sculpture and architecture. Yet the measured precision it allowed provided the basis for engineering and instrument-making, even the science of warfare. The mastery of perspective projection and the geometrical casting of shadows was central to it all.

The drawings of the architect Anton Hallmann — on show in the exhibition *Apoll im Labor* at the Berlin Museum of Medical History until 2 October — exemplify the heights reached by technical draftsmanship in the nineteenth century. Introduced to projective geometry at the Artillereschule in Hanover, Hallmann became a master of precise architectural representation, and was best known for his depiction of buildings of classical antiquity. His drawings range from the meticulous perspectival rendering of the whole and parts of buildings to extraordinary abstract exercises in which he set geometrical bodies in measured spaces under specific illumination.

In the example shown here, he scatters a series of bodies across a stage-like ground on which two torches provide what are taken to be point sources of light. The shadows are observed in full geometric projection, as they cast a complex ensemble of angular and conic contours across the space. The relative intensities of the two overlapping shadow systems are meticulously computed, with often surprising results.

Hallmann's link with science is his friendship with Emil Du Bois-Reymond, the pioneer of electrophysiology and inventor of

instruments for measuring biological forces.

There is an obvious connection through Hallmann's drawings of the mechanical system of bones and muscles in the human body, but the affinity goes deeper than a common subject. What they and other Berlin practitioners of the arts and sciences shared was a passion for classical clarity.

Du Bois-Reymond depicted himself in an engraving as a semi-nude classical god using his *Multiplikator* — a precision galvanometer — to measure his muscular electricity (see *Nature* **436**, 27; 2005). He designed his own instruments, which were constructed and operated with a love of form and space in which the beauty of pure mathematics was conjoined with the practical measurement of the forces of nature. This was euclidian mathematics embodied in human form, as exemplified in ancient sculpture.

The network of relationships extends to

Du Bois-Reymond's fellow members of the Berlin Physical Society, the psychologist Ernst Brücke and the physicist Hermann von Helmholtz, both of whom were keenly involved in the visual arts. Du Bois-Reymond had harboured ambitions to be a painter and taught anatomy to the students in Berlin's art academy. They all shared a vivid sense of the aesthetics of the powers of nature, as revealed through measurement and charted most potently on the curves of their graphs.

In Hallmann's primary field of architecture, the net extends to the classicist Karl Friedrich Schinkel, master builder and engineer, and his students at the Bauakademie, whose building he designed in a semi-industrial style in 1831. In the perspectival depiction of cityscapes, the net embraces Eduard Gärtner, master-painter of Berlin panoramas, and Johann Hummel, famed for his paintings' optical exactitude (*Nature* **395**, 649; 1998).

Within this Berlin nexus, whether you start with a physicist or a painter, you can connect to any other discipline in just two or three moves. No division into 'two cultures' here.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK.



NEWS & VIEWS



T. MORRISON/SOUTH AMERICAN PICTURES

The hottest hotspot of them all?
The tropical Andes is a hotspot under three defining criteria³.

BIODIVERSITY

Turning up the heat on hotspots

Hugh P. Possingham and Kerrie A. Wilson

Different measures are used to define concentrations of biodiversity — so-called 'hotspots'. More rigorous, global-scale analyses of how they compare will be essential for efficient resource allocation to conservation.

The variety of life on Earth is in rapid decline¹, and global spending on nature conservation is inadequate to arrest that decline². Consequently, resources for conservation must be allocated to secure the 'biggest bang for our buck'. In recognition of that need, scientists have identified biodiversity hotspots, where extraordinary concentrations of biodiversity exist, as defined by one or more metrics: number of species (species richness); number of species restricted to a particular area (endemic-species richness); and number of rare or threatened species. As well as using these three 'diversity metrics', we need to determine the seriousness of the threat to biodiversity hotspots from processes such as land-use changes³.

There is considerable controversy about which metric(s) to use to delineate hotspots, and the consequences of applying different metrics⁴. Using a newly compiled global database for the breeding distribution of all of the world's birds, Orme *et al.* (page 1016 of this issue)⁵ enter the fray. They have discovered an

alarming lack of congruence between hotspots defined using different metrics — so raising concern about the notion that species-richness hotspots should overlap with those identified using the other two diversity metrics⁶. It also calls into question the use of the hotspots principle in setting priorities for conservation.

The science of biodiversity hotspots is part of the emerging interdisciplinary field of conservation biogeography. The seminal definition of this field, by Whittaker *et al.*⁷, is "the application of biogeographical principles, theories, and analyses, being those concerned with the distributional dynamics of taxa individually and collectively, to problems concerning the conservation of biodiversity". The same authors⁷ also list the factors that reduce the value of such research; some of these illuminate the roots of past controversies.

One factor that can lead to spurious results is a lack of good data on the distribution of groups of species that are not taxonomically well defined. Orme *et al.*⁵ largely circumvent this problem by using birds, which have

well-known distributions and are relatively stable taxonomically. The spatial scale of the analysis is also critical, both with respect to its geographical extent and the resolution of data. The scales of Orme and colleagues' analysis are respectively the world and equal-area grids of approximately 1° latitude by 1° longitude; such resolution provides consistent and relatively fine-resolution data for a global analysis. Finally, there is a frustrating lack of theory to explain the reasons for the different patterns of biodiversity that have been identified. Orme *et al.* address this point by proposing mechanisms responsible for the observed geographical patterns of different aspects of biodiversity.

Compelling theories on why hotspots of different kinds occur where they do are beginning to emerge from analyses of data on the global and continental scales^{5,6}. Orme and colleagues' geographical analyses of the three diversity metrics provide further insights into the ecological, evolutionary and human effects that underlie the origin and maintenance of biodiversity as measured in these ways. They

argue that these effects are largely associated with large-scale topography. This supports the view that conservation assessment must consider not only contemporary factors, but also historical and ecological factors, such as the distribution of stable environments (which affects the extinction rate of species with small ranges) and the existence of geographical barriers and steep environmental gradients (which generate the diversity). The authors also argue that the distribution of threatened species is determined by an interaction of biological and human factors, and that the human factor helps to explain the lack of congruence between hotspots defined by the number of threatened species and those defined by the other two diversity metrics.

This lack of congruence highlights the potential inefficiencies that arise from using a single metric to delineate hotspots, and taking that to guide conservation efforts. Nevertheless, it is becoming clear that if resources are expended on endemic-species hotspots, they are likely to go a long way in protecting both species-richness and threatened-species hotspots. The endemism hotspots identified by Orme *et al.*⁵ (for example, the tropical Andes, pictured) contained a greater proportion of species richness than did the species-richness hotspots, and a greater proportion of threatened species than the threatened-species hotspots. Areas with large numbers of endemic species may also be of special significance in setting conservation priorities, because they may be areas of high past, and potentially future, speciation⁶.

All is not doom and gloom for the hotspots principle, however, as congruence between different studies using completely different methodologies is invariably high⁸. All 10 threatened-bird-species hotspots identified by Orme *et al.* are on the Conservation International list of hotspots⁹, which is based on plant endemic richness and habitat loss³. And only two of Orme and colleagues' 20 endemic-species hotspots are not on the list — one of these two, the Guyana Highlands of northern South America, is also a species-richness hotspot that may warrant further global attention.

Earlier this year, Hoekstra *et al.*¹⁰ added a new dimension to the debate over conservation priorities. They ignored species, and instead adopted a habitat-based approach¹¹. They show that the temperate grassland and Mediterranean biomes of the world are those most in need of urgent protection, and so counter the prevailing wisdom that conservation resources should be concentrated in tropical habitats. Analyses of taxonomic groups other than birds, and a marriage of species-based and habitat-based approaches, should go a long way to providing a robust vision of conservation priorities for the future.

The amount and quality of global data on biodiversity is increasing rapidly, and there will be a continued refinement of — possibly

even consensus about — the location of biodiversity hotspots. However, the cost of conservation action, which varies by several orders of magnitude from place to place², is an essential factor missing from this research agenda. If hotspots research is primarily an exercise in the study of spatial patterns of biodiversity and threats to biodiversity, costs are irrelevant. But if its real purpose is to guide resource allocation for conservation where time and money are constraints, we must urgently work to include economic and social factors. ■

Hugh P. Possingham and Kerrie A. Wilson are at the Ecology Centre, University of Queensland,

Brisbane, Queensland 4072, Australia.

e-mail: h.possingham@uq.edu.au

1. Dirzo, R. & Raven, P. H. *Annu. Rev. Environ. Res.* **28**, 137–167 (2003).
2. Balmford, A., Gaston, K. J., Blyth, S., James, A. & Kapos, V. *Proc. Natl Acad. Sci. USA* **100**, 1046–1050 (2003).
3. Myers, N. *et al. Nature* **403**, 853–858 (2000).
4. Mace, G. M. *et al. Nature* **405**, 393 (2000).
5. Orme, C. D. L. *et al. Nature* **436**, 1016–1019 (2005).
6. Jetz, W., Rahbek, C. & Colwell, R. K. *Ecol. Lett.* **7**, 1180–1191 (2004).
7. Whittaker, R. J. *et al. Divers. Distrib.* **11**, 3–23 (2005).
8. Myers, N. *BioScience* **53**, 916–917 (2003).
9. www.biodiversityhotspots.org/xp/Hotspots
10. Hoekstra, J. M. *et al. Ecol. Lett.* **8**, 23–29 (2005).
11. Morrison, J. C. *et al. BioScience* **51**, 933–938 (2001).

COSMOLOGY

Original questions

Martin Bojowald

The lack of a coherent quantum description of gravity has impeded our understanding of the physics that determined how the Universe began. A synthesis of recent ideas may take us a step farther back in time.

Among the deepest, borderline-philosophical questions in modern physics is that of the origin and formation of the Universe. Earlier attempts to formulate an answer that takes into account existing theories and observations have failed because of obstacles posed by gravity. Mulryne *et al.*¹, writing in *Physical Review D*, provide a 'loop quantum gravitational' model that successfully merges current ideas, and which may enable us to overcome such difficulties.

The most important feature to bear in mind when considering the origin of the Universe is the radiation that was released when the Universe became transparent to light, the so-called cosmic microwave background². Anisotropies in this radiation — slight variations in its temperature according to the direction in which you look at it — carry information on the distribution of matter at the time of its release. Through backward evolution of theoretical models of the Universe, we can garner an idea of what the initial seeds for any structure we observe in the cosmos might have been. The currently favoured models are inflationary models, and postulate an accelerated expansion of the early Universe at the time when the initial seeds were being sown.

The trouble with these models is that they require a state at which space is not just tiny, but has no size at all, and where the amount of energy stored becomes infinite — a situation impossible to deal with in the classical theory on which they rely. Mathematically, this is a 'singularity', where the main equations and concepts of a theoretical framework become inapplicable. Quite often, this state of zero size is speculatively identified as the 'initial' state of the Universe. However, it is simply ill-defined

in the theory of general relativity, which is our current best description of the nature of space and time.

Near a singularity, we reach the limits of current theory. At extremely small sizes and high energies, quantum effects are expected to be significant, so a quantum theory of gravity is needed. The required combination of general relativity and quantum theory has so far resisted consistent formulation. We can, however, attempt to apply some promising candidate theories to the early Universe. Loop quantum gravity^{3,4} is one such theory; it can deal with both strong gravity and a potentially vanishing space, and can be applied to cosmological situations in a framework known as loop quantum cosmology⁵. The theory gives rise to characteristic effects, such as the energy in matter in quantized space behaving differently, on small scales, from how it does in classical formulations⁶. To some degree, quantum space can be considered as analogous to a crystal, which, through its atomic structure, changes the propagation of light relative to that through a vacuum.

One characteristic consequence of these features of loop quantum cosmology is a repulsive contribution to the classical, attractive force of gravity. It is easy to imagine that this repulsion could prevent the total collapse of the Universe to zero size⁷, or even, when it is expanding, accelerate that expansion⁸. Combined with ideas of inflationary cosmology, the proposition has the ingredients of a well-defined and observationally viable model. Yet by itself it still does not explain the origin of the Universe.

One attempt at such an explanation is the emergent-Universe model^{9,10}. In the absence of additional information on the initial state of

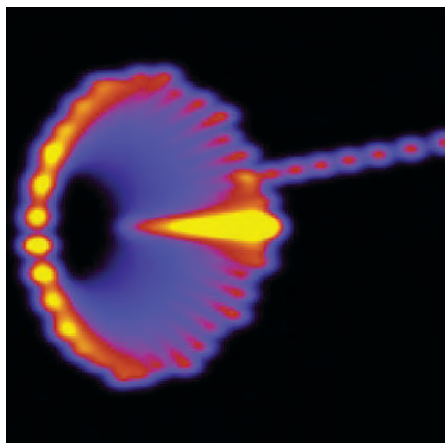


Figure 1 | Out of the loop. A small Universe initially cycles through different sizes, and eventually escapes to an inflationary era (line off to right). Colours represent how often the Universe reaches a certain size, growing from left to right; the position in the vertical direction is determined by the amount of expansion or contraction.

a system, it is most economical to assume that it was the simplest possible: for a physicist, this means the most highly symmetrical. Such a state would not have any structure in space or time, but would be homogeneous and static — an assumption already considered by Einstein. Classical examples of such states, called Einstein static spaces, do exist, provided that space is curved and closed.

Static solutions do not evolve, and so are clearly ill-suited as a model for the Universe. But by introducing a perturbation to a static solution, one can slightly change it and thereby start a more interesting history. Unfortunately, the classical solution is unstable: any disturbance grows rapidly, leaving little of the initial state behind. The insight of Mulryne and colleagues¹ is that quantum effects could supply all the necessary ingredients where classical solutions do not. Within the framework of loop quantum gravity, repulsion also implies static solutions at small size, but these — in contrast to the classical case — are stable.

According to the authors' model, perturbing such a state leads to small cycles of interchanging expansion and contraction. During this process, matter will evolve slowly, and the cycles will gradually change their behaviour. By itself, this perpetual recurrence and incremental change seems to lack the spark necessary for so momentous an event as the birth of the Universe. And indeed, Mulryne and colleagues identify one final theoretical ingredient that lights this spark: mediated through repulsive effects, potential energy is gradually pushed into the matter during its slow evolution. At the point when potential energy starts to dominate kinetic energy, the mundane cycling is broken by a sudden, dramatic inflationary explosion — the emergent Universe (Fig. 1).

Mulryne and colleagues thus supply a promising, well-defined picture of how the Universe with its complicated structure could

have emerged from a simple initial state. It is unlikely that the existence of any new observable effects will be postulated soon on the basis of this picture, although it does clarify several conceptual problems: the possibility of non-singular behaviour, for example, and the role of closed spaces. The basic effects interpreted as repulsion have been known as mathematical constructs for some time. But it is only when incorporated into cosmological models such as these, and models governing the physics of black holes, that we see how important quantum-gravitational effects can be, and how naturally they can fill in the gaps in our knowledge.

Work such as that of Mulryne *et al.*¹ gives strong support to general ideas of quantum gravity, although various models and effects must still be better justified and tied in more closely to a full theory. The virtue of cosmological investigations lies not only in their

being a supplier of basic ideas, but also in their guiding of developments and showing where, in so complex a theory as quantum gravity, one should look for interesting effects. ■

Martin Bojowald is at the Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Am Mühlenberg 1, D-14476 Golm, Germany. e-mail: mabo@aei.mpg.de

1. Mulryne, D. J., Tavakol, R., Lidsey, J. E. & Ellis, G. F. R. *Phys. Rev. D* **71**, 123512 (2005).
2. Spergel, D. N. *et al. Astrophys. J. Suppl.* **148**, 175–194 (2003).
3. Ashtekar, A. & Lewandowski, J. *Class. Quantum Grav.* **21**, R53–R152 (2004).
4. Rovelli, C. *Quantum Gravity* (Cambridge Univ. Press, 2004).
5. Bojowald, M. in *100 Years of Relativity* (ed. Ashtekar, A.) (World Scientific, Singapore, in the press); preprint at www.arxiv.org/gr-qc/0505057.
6. Thiemann, T. *Class. Quantum Grav.* **15**, 1281–1314 (1998).
7. Singh, P. & Toporensky, A. *Phys. Rev. D* **69**, 104008 (2004).
8. Bojowald, M. *Phys. Rev. Lett.* **89**, 261301 (2002).
9. Ellis, G. F. R. & Maartens, R. *Class. Quantum Grav.* **21**, 223–232 (2004);
10. Rebhan, E. *Astron. Astrophys.* **353**, 1–9 (2000).

ATMOSPHERIC CHEMISTRY

Natural bleach under scrutiny

Patrick Jöckel and Carl A. M. Brenninkmeijer

Cosmic rays produce carbon-14, which enters Earth's carbon cycle after being oxidized. It is of great service to atmospheric chemists in providing a way of tracking the degree to which the atmosphere keeps itself clean.

As Martin Manning and colleagues report on page 1001 of this issue¹, changes in the amount of hydroxyl (OH) radicals in Earth's atmosphere can be tracked by analysing time-series measurements of naturally produced carbon monoxide containing radiocarbon (¹⁴CO). This is no mean feat and is of considerable significance — OH is the chief oxidant in Earth's atmosphere, and as such acts as a natural bleaching agent. Atmospheric chemists have been struggling to estimate how much OH there is, and how much it varies in concentration in space and time: Manning and colleagues' approach constitutes a big step forward.

The self-cleansing capacity of Earth's atmosphere is remarkable. Every year, roughly half-a-billion tonnes of methane (CH₄) and 2.5 billion tonnes of CO are removed from the troposphere by chemical reaction. (The troposphere is the lowermost layer of the atmosphere; it is the site of the machinery that creates weather, and extends 10–15 km above Earth's surface.) This miracle of self-cleansing occurs even though CH₄, CO and several other reduced gases do not react at any significant rate with the atmosphere's major oxidant, molecular oxygen (O₂), or the rarer but more powerful ozone (O₃).

The true cause was not discovered until 1971, when it was recognized² that even in remote regions, far away from photochemical

smog, active atmospheric chemistry occurs. The breakdown of O₃ by ultraviolet sunlight produces excited oxygen radicals. Some reform into O₃. But others retain enough energy to split water molecules and create OH radicals. These are stable but highly reactive, and constitute the troposphere's bleaching agent. Thanks to OH, the lifetime of CH₄ (a greenhouse gas) is kept below ten years, whereas on average a CO molecule perishes in a matter of months in the reaction CO + OH → CO₂ + H.

It is this last reaction, but using ¹⁴CO, that Manning *et al.*¹ have exploited to estimate levels of OH. The lifetime of OH is merely one second, making direct measurements technically demanding. Its extreme reactivity not only implies low abundances, at an average level of 1 million radicals per cm³ (below part-per-trillion levels), but also great variability in its concentration — from night to day, from cloudy sky to clear sky, from summer to winter, and depending on latitude.

But ¹⁴CO, which originates from ¹⁴C produced by cosmic rays (Fig. 1), is an excellent natural tracer for tracking OH. The principle of this indirect approach was first outlined by Bernard Weinstock³: when the oxidative capacity of the atmosphere falls, with fewer OH radicals present, ¹⁴CO levels can rise because the rate of removal of ¹⁴CO — via oxidation by OH to ¹⁴CO₂ — is lower. Rates of production and destruction are assumed to be in equilibrium.

The product $^{14}\text{CO}_2$ is, of course, well known in the environmental sciences: following its uptake by plants and subsequent entry into the food chain, its radioactive decay provides the basis for radiocarbon dating.

Manning and colleagues have analysed 13 years of ^{14}CO measurements at Baring Head, New Zealand, together with similar data from Antarctica and ship cruises. After correcting the time series for the large modulation of ^{14}C production caused by the 11-year solar cycle, residual variations in ^{14}CO remain. As the authors argue, two instances of higher ^{14}CO can only have been caused by short-term reductions in OH, and the coincidence with known atmospheric changes confirms their hypothesis. Their work clearly shows one of the advantages of using ^{14}CO for tracking OH. Because of its short lifetime, ^{14}CO is sensitive to rapid atmospheric changes such as those that occur after major volcanic eruptions or large-scale episodes of biomass burning related to El Niño climatic events. It is also notable that the short lifetime of ^{14}CO enabled the authors to consider the remote Southern Hemisphere as a fairly self-contained atmospheric 'laboratory' for testing its use.

Computer models of atmospheric transport and chemistry can generate a fairly detailed picture of OH distribution. Typically, the maximum values occur in the tropics, as might be expected: it is here that atmospheric chemistry is at its most active because of the intense solar radiation and high amounts of water vapour. Verifying the model picture is another matter, and a previous approach that has been repeatedly applied is based on methyl chloroform, which also offers an indirect way of estimating OH levels. Careful measurement of this chlorinated industrial chemical at several locations, and calculation of emissions from manufacturers' data, have shown that OH has apparently undergone surprisingly large changes over the past decades⁴. Yet this and related findings have been controversial^{5,6} because of uncertainties about the actual rates of emissions. Moreover, the production of methyl chloroform has been phased out, and — thanks to OH — it is disappearing from the

atmosphere. So this is not a tracer that can be used in the long term.

By contrast, ^{14}CO is produced naturally and largely independently of human activity. It should become the principal diagnostic tool for monitoring the oxidative capacity of the atmosphere now and in decades to come. This tracer is a cosmic dowry for atmospheric chemists — Manning *et al.* have made a strong case for them to accept it with gratitude. ■

Patrick Jöckel and Carl A. M. Brenninkmeijer are in the Department of Atmospheric Chemistry, Max Planck Institute for Chemistry, PO Box 3060, 55020 Mainz, Germany.

e-mail: joeckel@mpch-mainz.mpg.de

1. Manning, M. R., Lowe, D. C., Moss, R. C., Bodeker, G. E. & Allan, W. *Nature* **436**, 1001–1004 (2005).
2. Levy, H. *Science* **173**, 141–143 (1971).
3. Weinstock, B. *Science* **166**, 224–225 (1969).
4. Prinn, R. G. *et al.* *Science* **292**, 1882–1888 (2001).
5. Spivakovsky, C. M. *J. Geophys. Res.* **96**, 17395–17398 (1991).
6. Krol, M. & Lelieveld, J. *J. Geophys. Res.* **108**, 10.1029/2002JD002423 (2003).

CELL BIOLOGY

Shaggy mouse tales

Elizabeth H. Blackburn

First impressions can be misleading. The enzyme telomerase has been well studied because of its initial association with cell ageing processes and cancer — but it now seems that this is not all it can do.

The way in which a biological entity is first identified can limit perceptions of the full range of its functions. The telomerase enzyme, for instance, was originally discovered on the basis of its vital ability to lengthen telomeres — the stretches of non-coding DNA at the ends of chromosomes. If it is too short, a telomere loses the ability to maintain a protective structure at the end of the chromosome, and such shortened telomeres can signal to cells to cease multiplying, in a process called cellular senescence. In this issue, Sarin *et al.* (page 1048)¹ report provocative evidence that telomerase does more than merely synthesize DNA at chromosome ends: a key subunit of the enzyme stimulates the proliferation of mouse hair-follicle stem cells, generating shaggy mice. Strikingly, this occurs independently of the DNA-synthesis capacity of telomerase.

Telomerase adds DNA to the tips of the chromosomes to replenish the telomeres. This DNA would otherwise dwindle away as cells multiply, owing to incomplete replication of the chromosomal DNA and to enzymes nibbling away at the DNA end regions. In normal human cells, telomerase is highly regulated, and its efficiency depends, among other things, on the cell type². The enzyme is present in very low amounts in most cultured human primary cells (that is, those grown directly from biopsies)³, so their telomeres gradually

shorten and the cells senesce. If such cultured cells are made to express excessive amounts of normal telomerase, the telomeres elongate and overcome cellular senescence, so the cells continue to multiply.

Telomerase is actually a complex of molecules, and its DNA-synthesis function requires a collaboration between two core components, a protein subunit called TERT and an RNA component called TERC in mice. To explore the function of TERT, Sarin and colleagues made a mouse strain that contained an extra TERT gene; the original TERT gene was left intact. The extra gene had a control element so that it could be turned on or off at will in the live animal. When switched on, the gene produced large amounts of TERT protein in all cells of the animal's body.

The authors found that when TERT was overexpressed in this way the mice were very furry, and that this was due to increased proliferation of hair-follicle stem cells. A similar effect was independently reported recently for TERT overexpressed only in mouse skin⁴. These stem cells produce the mature follicles from which hair grows, and the extent of their proliferation controls fur growth. The authors next timed the TERT overexpression to occur at specific periods in the cycling of the hair follicles between their active (anagen) and resting (telogen) stages. They found that the

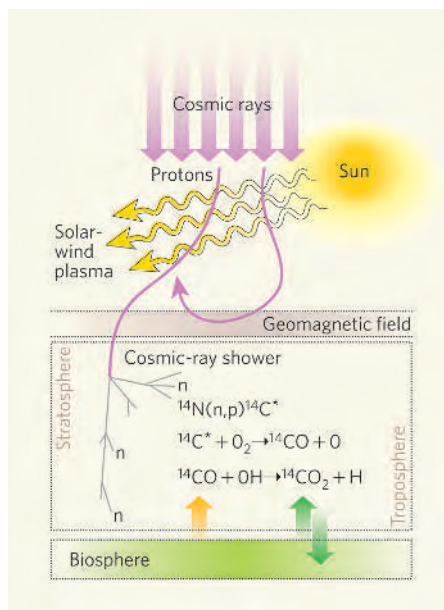


Figure 1 | Production of ^{14}C , and the OH connection. Cosmic rays (mainly protons) are scattered and decelerated by the solar-wind plasma. But some retain enough energy to penetrate the geomagnetic field and enter Earth's atmosphere as showers of cosmic rays. Nearly all of the neutrons (n) produced in this process are captured by nitrogen nuclei, which lose a proton and form atoms of excited radiocarbon, $^{14}\text{C}^*$, which then mainly react to form ^{14}CO . The subsequent oxidation of ^{14}CO by OH radicals to $^{14}\text{CO}_2$ is the key reaction that allows Manning and colleagues' measurements¹ of ^{14}CO to be used in estimating variations of OH concentrations in the atmosphere. This cosmogenic source accounts for about 75% of the ^{14}CO in the atmosphere. The remainder is recycled from the biosphere by biomass burning and oxidation of volatile organic carbons.

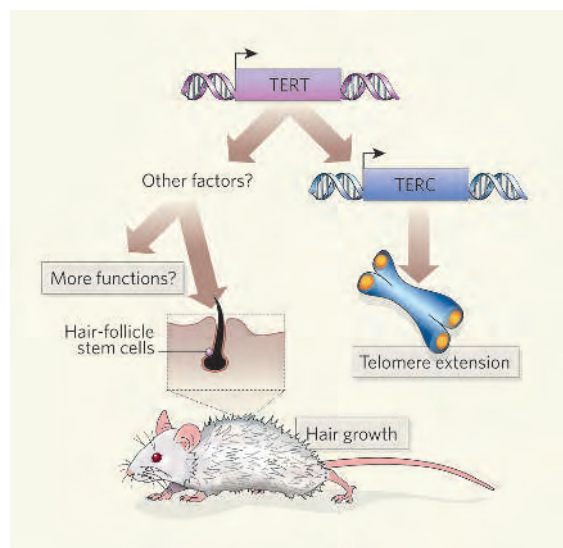


Figure 1 | Telomerase is more than it seemed. The telomerase enzyme was first discovered based on its ability to add DNA to the telomere regions at the ends of chromosomes. It works as a complex with two core components: the protein TERT and an RNA component called TERC in mice. However, Sarin *et al.*¹ find that TERT can also stimulate the proliferation of hair-follicle stem cells and hair growth — even in the absence of TERC. Perhaps other functions of telomerase have yet to be found.

overexpressed TERT could reawaken telogen hair follicles, causing them to move into the anagen phase and promoting hair growth. Most remarkably, the prolific hair growth was apparent even when the TERT gene was overexpressed in a mouse strain that lacked the RNA component of telomerase (Fig. 1), showing that the effects on hair growth are independent of the DNA-synthesis function.

How might the core-protein subunit of telomerase make mice shaggy in the complete absence of its function of telomeric DNA synthesis? The answer is unclear. Evidence that telomerase has additional biological roles began to surface in the late 1990s, through experiments on the enzyme in yeast and human cells⁵. Certain partially active mutant forms of telomerase were found to overcome cellular senescence despite massive shortening of the telomeres; the cells continued to multiply even though their telomeres were shorter than those normally seen in senescent cells⁵. In addition, overexpressing normal telomerase in cultured human primary cells changed the patterns of gene expression across the whole genome, whereas cell growth rates and overall telomere length did not change noticeably⁶. Moreover, reducing even the small amount of functional telomerase in normal human fibroblasts (connective-tissue cells) accelerates their senescence⁷. Even before their telomeres shorten, this quickly compromises normal protective cellular responses to agents that damage DNA³.

Human cancer cells commonly have high telomerase activity, although their telomeres are typically short. However, blocking telomerase production rapidly inhibits cancer-cell growth without telomere shortening, and alters the cells' gene-expression profile in a distinctive fashion that may be associated with diminished cancer progression⁸. Indeed, an *in vivo* model of skin cancer showed that inhibiting production of the telomerase RNA reduced metastasis⁹. Although certain cells lacking telomerase can maintain their telomeres by an alternative mechanism,

which is independent of telomerase, the tumour-generating capacity of such cells is less than that of cells expressing telomerase¹⁰. Conversely, overexpressing TERT in mice promotes early progression of skin cancer⁴.

How all these effects are mediated, and whether or how they involve telomerase activity on or off the telomere, are unknown. TERT is known to form at least one complex without telomerase RNA — yeast TERT binds to a protein called PinX1 in a manner that is mutually exclusive of its binding to telomerase RNA¹¹.

So perhaps TERT forms other complexes with as-yet-unidentified partners.

In ancient Egypt, men smeared their pates with hippopotamus fat in a desperate bid to stave off baldness¹². Is telomerase the new hippopotamus fat? Probably not. But this enzyme is already known to be vital in sustaining tissues in health and disease, and we should look beyond its eponymous function to understand the full spectrum of its potential roles. ■ Elizabeth H. Blackburn is in the Department of Biochemistry, University of California, San Francisco, 600 16th Street, San Francisco, California 94143-2200, USA. e-mail: telomer@itsa.ucsf.edu

1. Sarin, K. Y. *et al.* *Nature* **436**, 1048–1052 (2005).
2. Epel, E. S. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 17312–17315 (2004).
3. Masutomi, K. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 8222–8227 (2005).
4. Flores, I., Cayuela, M. L. & Blasco, M. A. *Science advance online publication* 21 July 2005 (doi:10.1126/science.1115025).
5. Blackburn, E. H. *Nature* **408**, 53–56 (2000).
6. Smith, L. L., Coller, H. A. & Roberts, J. M. *Nature Cell Biol.* **5**, 474–479 (2003).
7. Kim, M. M. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 7982–7987 (2001).
8. Li, S., Crothers, J., Haqq, C. M. & Blackburn, E. H. *J. Biol. Chem.* **280**, 23709–23717 (2005).
9. Nosrati, M. *et al.* *Clin. Cancer Res.* **10**, 4983–4990 (2004).
10. Stewart, S. A. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 12606–12611 (2002).
11. Lin, J. & Blackburn, E. H. *Genes Dev.* **18**, 387–396 (2004).
12. Ebers papyrus ca. 1550 BC Univ. Leipzig, Spec. Collections Dept.

ASTROPHYSICS

Swift progress

Dieter H. Hartmann

The agile, choreographed response of the Swift satellite to γ -ray bursts tests models to an unprecedented degree. Results from two recent long bursts suggest that the models are good, but require some tweaking.

The Swift satellite¹ is NASA's latest tool for investigating the mysterious phenomenon of γ -ray bursts, or GRBs. These intense bursts of high-frequency γ -ray and X-ray radiation were discovered four decades ago²; since Swift's launch in November 2004, one such burst has activated its detectors every few days. On page 985 of this issue, Tagliaferri *et al.*³ present observations from Swift that suggest strongly that the 'prompt' emission of the burst and its so-called afterglow are separate radiation components. This is consistent with models in which the prompt phase of γ -radiation, lasting a few to a few hundred seconds, results from internal shocks within the medium that produces them, whereas the longer-lasting afterglow, at lower frequencies, is the outcome of interactions between ejected and surrounding material.

The durations of GRBs, together with their spectral properties, suggest a classification into short and long bursts⁴. How far away both

classes originate was unknown until a distance scale was established, at first on a statistical basis⁵ by comparing the directions of GRBs with models of their possible distribution in the Galaxy. This was followed by direct determination through the discovery of X-ray afterglows⁶ and the measurement of redshifts through optical line spectroscopy⁷. Separate models for the origins of short and long GRBs also emerged. Short GRBs are believed to be associated with the merger of compact binaries: a pair of neutron stars, for example. This explanation is not yet certain, and the Swift mission is designed to clarify the nature of the short bursts.

For long GRBs, the favoured 'collapsar' model⁸ starts from the idea of a rapidly rotating, massive star that has undergone extreme gravitational collapse and formed a central black hole. Complex processes extract some of the gravitational binding energy from the disk of material that forms around this black hole;



50 YEARS AGO

Life, the Great Adventure. By Jean Rostand. (Discussions with Paul Bodin.) — An admirable guide for those who wish to explore some of the broader paths in the field of current biology... With Rostand's general point of view few modern readers would quarrel, although his explanation of reasons for superfluous hair in the male is nonsense... These, however, are trivial criticisms of a book which could reach many readers. Many of them will be shocked to learn that... in a country like France, whose people are so proud of their ability to behave rationally, there are still about 3,460 astrologer-palmists in Paris alone... One of the lessons from this book is that the mass of people are still prepared to believe anything, particularly if the new cult is served persuasively with the right amount of scientific jargon. Rostand and Bodin deserve praise for their efforts to remove some of the superstitions of 1955. From *Nature* 20 August 1955.

100 YEARS AGO

The *Academy* directs attention to a curious poetical tribute — composed by a French mathematician — to Archimedes, referring to the evaluation of π , which, set out in thirty places of decimals, is 3.141592653589793238462643 383279. It will be observed that each of the thirty-one words in this quatrain contains the number of letters corresponding with the successive numbers in the numerical expression: — *Que j'aime à faire apprendre un nombre utile aux sages/Immortel Archimède, artiste ingénieur/! Qui de ton jugement peut priser la valeur?/Pour moi ton problème eut de pareils avantages.*

The *Frankfurter Zeitung*... adds a similar effort emanating from a German poet and geometrician: — *Dir, o Held, o alter Philosoph, Du Riesen-Genie!/Wie viele Tausende bewundern Geister/himmlisch wie Du und göttlich!/Noch reiner in Aeonen/wird das uns strahlen/wie im lichten Morgenrot!*

The *Academy* asks for English parallels to these efforts. From *Nature* 17 August 1905.

directed along the axis of rotation, this energy breaks through the surface of the star at almost the speed of light. This high-energy radiation shockwave is sent out along a narrow jet and can be spotted by detectors — if the jet is pointing the right way.

According to the collapsar model, the prompt energetic structure of a long GRB is the result of dissipation by internal, relativistic shocks, which may last seconds or minutes, at a radius of about 10^{14} cm from the centre of the collapsed star⁹. At this point, much of the energy remains in storage as kinetic energy and is tapped later¹⁰, at a radius of about 10^{16} cm, when the outflow collides with the interstellar medium or the particle wind from the progenitor star. This two-stage model, known as the internal-external model¹¹, predicts a transition from an erratic, prompt γ -ray outburst to a smoother X-ray afterglow (Fig. 1). Late bumps in the light curves that are observed from long GRBs, as well as directly observed spectral features, are reminiscent of features from stellar explosions — supernovae — and provide evidence in support of the collapsar model.

Tagliaferri and colleagues observed³ two long bursts, GRB050126 and GRB050219a, which lasted about 20–30 seconds and were detected initially by Swift's Burst Alert Telescope (BAT) at γ -ray frequencies. True to its name, Swift rapidly — in 129 seconds in the first case and 87 seconds in the second — slewed its X-Ray Telescope (XRT) into the line of the burst to allow unprecedented monitoring, at low X-ray frequencies, of the afterglow. The Swift data show that, as the prompt emission fades, an initial afterglow appears that declines very rapidly over a few hundred seconds. This early afterglow is, however, distinct from the later afterglow familiar from pre-Swift data — with the high sensitivity and excellent time coverage of Swift, we are now witnessing the X-ray light curves in the transition period from prompt emission to afterglow.

Although the Swift data broadly support the notion of two components to GRB emission, the decay of the burst's prompt brightness within the first few minutes is unexpectedly steep, falling with the inverse cube of time, t^{-3} . How should we interpret this? The authors cite several possibilities³, but most of these are undermined by the 'non-thermal' spectrum of the radiation that follows a power law, rather than the 'blackbody' form assumed in the models. The most intriguing option that would provide such a rapid decay is 'inverse Compton' scattering of low-energy photons by high-energy electrons in the reverse shock of the burst. (Blast waves lead to two shocks, one propagating ahead into the external medium, and another propagating back through the ejected material of the jet.) If the inverse Compton mechanism dominates — which would be true if the outflow contained a very large abundance of electron-positron pairs — prompt 'optical' flashes that are also not observed might be suppressed¹². Robotic

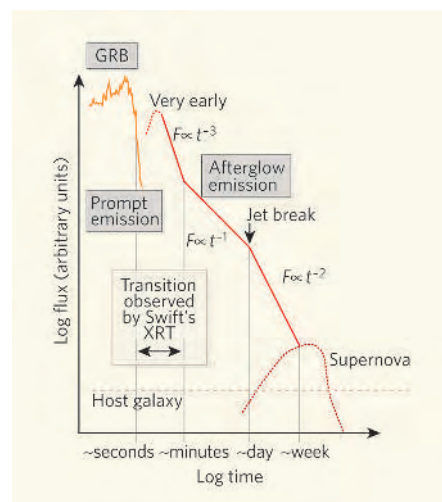


Figure 1 | Staged burst. The flux F of a γ -ray burst (GRB) over time t , as recorded by the Swift satellite and reported by Tagliaferri *et al.*¹. The prompt γ -ray emission usually lasts for a only few seconds, and is believed to originate from shocks within the relativistic jet. The longer-lasting afterglow, in the X-ray to optical and infrared bands and typically outshining the GRB's host galaxy, decays following a power law, with the flux falling as t^{-1} during the first 24 hours, and faster (t^{-2}) thereafter. In the transition region between prompt emission and afterglow, the gradient is even steeper (t^{-3}). A bump in the light curves sometimes observed a few weeks after the trigger is predicted by the collapsar model for long GRBs, and is believed to be light from an associated supernova.

telescopes on the ground with swift responses should soon constrain these speculations.

The standard model used to explain GRB afterglows is known as the fireball model and is based on a few main ingredients. First, the extreme brightness of the GRB, in conjunction with the large distances involved, implies a huge energy release. Second, the brevity of the bursts indicates a small volume in which this energy is released. Taken together, these seem to imply the existence of a dense, opaque fireball driving the burst to relativistic speeds.

If there really were a fireball, this would imply emission in all directions, and the observed GRB energies would be difficult to explain within the confines of conventional astrophysics. But commonly observed breaks in the afterglow suggest that the outflow is indeed jetted and is thereby limited to a small solid angle, reducing the energy required. After correcting for this effect, the bursts seem to behave approximately as 'standard candles'¹³ — objects whose observed brightness depends only on how far away they are — making them useful measures of cosmological distance¹⁴.

Continued monitoring of the sky with Swift, with other detectors such as HETE-2 and INTEGRAL, and with the future missions GLAST-GBM, Agile and EXIST, will further expand our understanding of GRBs. They will remain a source of amazement and wonder for some time to come.

Dieter H. Hartmann is in the Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29634-0978, USA. e-mail: hdieter@clemson.edu

1. Gehrels, N. *et al.* *Astrophys. J.* **611**, 1005–1020 (2004).
2. Klebesadel, R. W., Strong, I. B. & Olson, R. A. *Astrophys. J.* **182**, L85–L88 (1973).
3. Tagliaferri, G. *et al.* *Nature* **436**, 985–988 (2005).
4. Kouveliotou, C. *et al.* *Astrophys. J.* **413**, L101–L104 (1993).
5. Meegan, C. *et al.* *Nature* **355**, 143–145 (1992).

6. Costa, E. *et al.* *Nature* **387**, 783–785 (1997).
7. Metzger, M. *Nature* **387**, 878–880 (1997).
8. MacFadyen, A. I. & Woosley, S. E. *Astrophys. J.* **524**, 262–289 (1999).
9. Rees, M. & Meszaros, P. *Astrophys. J.* **430**, L93–L96 (1994).
10. Rees, M. & Meszaros, P. *Mon. Not. R. Astron. Soc.* **258**, 41p–43p (1992).
11. Piran, T. *Rev. Mod. Phys.* **76**, 1143–1210 (2004).
12. Kobayashi, S. *et al.* *Astrophys. J. Lett.* (submitted); preprint at www.arxiv.org/astro-ph/0506157 (2005).
13. Bloom, J. S., Frail, D. & Kulkarni, S. R. *Astrophys. J.* **594**, 674–683 (2003).
14. Ghirlanda, G. *et al.* *Astrophys. J.* **613**, L13–L16 (2004).

CARDIOLOGY

Rips repaired

Richard A. Steinhardt

In Duchenne muscular dystrophy, muscle cells die as a result of suffering many tiny membrane ruptures. A compound that increases membrane resealing can protect heart muscle cells from these effects.

Degenerative muscle diseases are usually the result of defects in the cytoskeleton, the cell's internal scaffolding. This structure normally shields muscle cells from mechanical stress, protecting their membranes from tearing. People with Duchenne muscular dystrophy lack the protein dystrophin, which is an essential link in the complex of proteins that connect the cytoskeleton in the cell to a matrix of extracellular proteins. These buttresses across the membrane protect it from mechanical stress. Consequently, Duchenne dystrophic cells tend to get more rips in their membranes than do normal muscle cells. In this issue, Yasuda *et al.* (page 1025)¹ show that the addition of a chemical that aids membrane resealing can preserve dystrophic heart muscle cells from such damage, and might help to maintain heart function *in vivo*.

Small rips in muscle cell membranes are normal, because these cells are under continuous mechanical stress. But in dystrophic muscle cells the frequency of these microdisruptions is greatly increased because their cytoskeleton cannot shield the membrane as effectively. Both normal and Duchenne dystrophic muscle cells can repair these tiny ruptures quickly, but not before calcium ions have flooded into the cell. This causes an increased calcium-dependent breakdown of proteins, which in turn results in a persistent calcium influx in the vicinity of the resealed wound. Over time there is a gradual loss of calcium homeostasis, further activation of protein degradation, and eventual cell death². As the dystrophic cells suffer many more rips than normal cells, cell death occurs more quickly.

Studies of Duchenne dystrophic cells have tended to concentrate on skeletal muscle rather than heart muscle, which under certain circumstances can be physiologically quite different. However, at least 15% of Duchenne sufferers die from heart failure. Yasuda *et al.*¹

have now analysed dystrophic heart tissue, and they confirm that individual dystrophic heart muscle cells also show more stretch-mediated damage to their membrane than do normal cells. Just as in skeletal muscle cells, this is followed by loss of normal calcium homeostasis and cell death.

The authors next examined the effect of a surfactant called poloxamer 188 (also known as Pluronic F68) on the dystrophic cells. This compound has been used in cell culture since the early 1960s to minimize mechanical damage from aeration and stirring of cell suspensions³. It is now clear that poloxamer 188 acts by helping to seal up tears in the cell membrane. Cell membranes are under considerable tension from their attachment to the cytoskeleton, and this tension is sufficient to prevent passive resealing of rips in the lipid bilayer. Instead, an active process called exocytosis adds membrane by fusion with small membrane-bounded vesicles from within the cell to lower the membrane tension and close the break^{4,5}. Poloxamer 188 greatly accelerates resealing, even in the absence of exocytosis⁶ — apparently by lowering the tension from cytoskeletal attachments at the membrane^{5,6}. This compound can insert directly into lipid monolayers⁷, and a similar process probably underlies its capacity to lower the tension in the bilayer lipid membrane.

Yasuda *et al.*¹ found that adding poloxamer 188 to dystrophic cardiac cells in culture mitigates any stress-induced injuries to the membrane. To test whether the effect is significant *in vivo*, they used a mouse model of muscular dystrophy, subjecting the mice to a chemical-stress protocol that would usually cause heart failure. Strikingly, injecting poloxamer 188 into the mice before the start of the stress protocol blocked the heart failure. So membrane fragility, membrane disruption and the resulting abnormal calcium homeostasis found in

dystrophic cardiac muscle can be corrected by an agent that rapidly reseals cell membranes. The authors propose that these results will have therapeutic implications for the treatment of muscular dystrophies.

A note of caution should be raised, however. Yasuda *et al.* demonstrate clear beneficial effects of adding poloxamer 188 to dystrophic heart muscle in the concentration range 0.5–1.25 mg per ml. But it is already known that the highest, most effective concentration is not well tolerated by healthy human volunteers given the compound for more than 44 hours⁸. The lowest dose, which is tolerated for up to 72 hours in humans, may well have serious consequences if taken continuously over a lifetime. Lifelong treatment would be necessary because poloxamer 188 is useful only if it is present before the break in the membrane occurs. Only then can it speed the resealing of cell membrane, bypassing the normal process⁶. Nevertheless, although poloxamer 188 is not ready as a therapy for lifelong conditions, the results of Yasuda and colleagues do imply that further research on membrane sealants could be very useful, especially if harmful side effects could be avoided, or if brief treatment periods were found to be of lasting benefit.

There is another application, however, which might be of immediate benefit to all. Major surgery carries with it a significant risk of oxygen deprivation that can lead to damage to the patient's heart or brain, and death. It has been known for some time that using an agent that artificially reseals the membranes of oxygen-deprived heart cells will prevent cell death for at least 24 hours after oxygen withdrawal⁹. That study⁹ used targeted liposomes (artificial lipid vesicles) to reseal the membranes, which is not practical as a medical procedure. But, given that damaged cells are very rapidly resealed in the presence of poloxamer 188 (ref. 6), and that this occurs *in vivo*¹, perhaps poloxamer 188 should be administered during high-risk surgery, at effective doses that are tolerated for 24 hours. The savings in terms of lives and the use of critical hospital resources might well be significant. ■

Richard A. Steinhardt is in the Department of Molecular and Cell Biology, University of California, Berkeley, California 94720-3200, USA. e-mail: ricksteinhardt@berkeley.edu

1. Yasuda, S. *et al.* *Nature* **436**, 1025–1029 (2005).
2. Alderton, J. M. & Steinhardt, R. A. *Trends Cardiovasc. Med.* **10**, 268–272 (2000).
3. Runyan, W. S. & Geyer, R. P. *Proc. Soc. Exp. Biol. Med.* **112**, 1027–1030 (1963).
4. McNeil, P. L. & Kirchhausen, T. *Nature Rev. Mol. Cell Biol.* **6**, 499–505 (2005).
5. Togo, T., Krasieva, T. B. & Steinhardt, R. A. *Mol. Biol. Cell* **11**, 4339–4346 (2000).
6. Togo, T., Alderton, J. M. & Steinhardt, R. A. *J. Cell Sci.* **112**, 719–731 (1999).
7. Maskarinec, S. A., Hannig, J., Lee, R. C. & Lee, K. Y. C. *Biophys. J.* **82**, 1453–1459 (2002).
8. Jewell, R. C., Khor, S. P., Kisor, D. F., LaCroix, K. A. & Wargin, W. A. *J. Pharm. Sci.* **86**, 808–812 (1997).
9. Khaw, B. A., Torchilin, V. P., Vural, I. & Narula, J. *Nature Med.* **1**, 1195–1198 (1995).

BRIEF COMMUNICATIONS

Lion attacks on humans in Tanzania

Understanding the timing and distribution of attacks on rural communities will help to prevent them.

Large carnivores inspire opposition to conservation efforts^{1,2} owing to their impact on livestock^{3–5} and human safety^{6,7}. Here we analyse the pattern of lion attacks over the past 15 years on humans in Tanzania, which has the largest population of lions in Africa^{8,9}, and find that they have killed more than 563 Tanzanians since 1990 and injured at least 308. Attacks have increased dramatically during this time: they peak at harvest time each year and are most frequent in areas with few prey apart from bush pigs (*Potamochoerus larvatus*), the most common nocturnal crop pest. Our findings provide an important starting point for devising strategies to reduce the risk to rural Tanzanians of lion attacks.

Figure 1a shows the reported number of attacks on people across Tanzania from January 1990 to September 2004 (for details of data collection, see supplementary information). More than 45% of all reported cases occurred in just six coastal districts in the southern half of the country, the area with the worst reputation for man-eating lions^{10,11}.

The total number of cases has increased strikingly since 1990 (Fig. 1b), probably because Tanzania's human population has risen from 23.1 million in 1988 to 34.6 million in 2002, with an associated loss of lion prey outside the protected areas¹². Nearly 39% of human attacks occur in the harvest season dur-

ing March–May ($n = 733$ of known date, $\chi^2 = 100.4$, $P < 0.001$). More than 18% of 538 victims of known age were younger than 10 years old; 69% of older victims were men, presumably because men are more likely to tend cattle, forage for bush-meat, walk alone at night and retaliate against man-eaters and cattle-killers with nets and spears, although women are also attacked in their homes and in fields (Fig. 1c). Lions pull people out of bed, attack nursing mothers, and catch children playing outside. Most rural houses have thatched roofs and many have thatched walls, so lions can force their way inside, and toilets are outside.

More than 27% of attacks occur in the fields (Fig. 1c), usually when people are sleeping in makeshift huts while protecting their crops from bush pigs (Fig. 2). Several interviewees specifically mentioned that lions entered their villages or fields in pursuit of bush pigs, and even reported tolerating lions because they helped to control bush-pig numbers.

Lions have been eradicated from many parts of the country, so our analysis is restricted to districts with at least one attack in the past 15 years. Attacks are most common in districts with the lowest abundance of natural prey (as estimated by the proportion of interviewees reporting kudu, zebra, hartebeest, dik-dik or impala; spatial covariance model: $P = 0.0162$, $n = 17$) and with the largest numbers of bush



Figure 2 | Makeshift hut in Rufiji district. Farmers typically sleep alone in these structures to protect their crops from nocturnal bush pigs during the harvest season, when the risk of attack by lions is highest. Lions force their way inside to pull out their victims.

pigs (spatial covariance model: $P = 0.0150$, $n = 17$). Multiple regression analysis reveals that nearly half the variance in attacks per district can be explained by these two factors (adjusted $r^2 = 0.45$, $P = 0.0059$, $n = 17$); there was no additional effect of human population density, cattle density, percentage of land cover devoted to agriculture, or proximity to a protected area.

Our analysis reveals a conspicuous outlier: Kilosa District had fewer lion attacks than expected from prey scarcity and bush-pig abundance. However, Kilosa had suffered high numbers of lion attacks until people were relocated by the Tanzanian government in 1992. If Kilosa is removed from the analysis, prey scarcity and bush-pig abundance account for 76% of the variance in lion attacks per district ($P < 0.0001$, $n = 16$).

Rising human populations mean that the option of relocating people to reduce the number of lion attacks is no longer feasible. It would also be impractical to increase natural prey populations in agricultural areas. As bush pigs are the most likely maintenance diet of lions in highly disturbed agricultural areas, controlling their numbers might be the best strategy to decrease the lions' attraction to populated areas and would reduce the need for village farmers to sleep in their fields (Fig. 2).

Human population growth has led to encroachment into wildlife areas and depletion of natural prey. However, conservation attempts to sustain viable populations of African lions place the lives and livelihoods of rural people at risk in one of the poorest

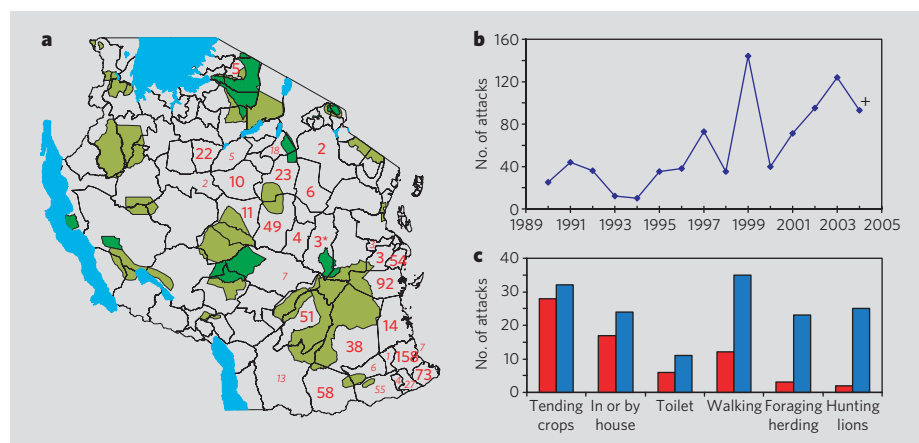


Figure 1 | Analysis of the number of lion attacks on humans. **a**, Map of Tanzania showing the distribution of 815 lion attacks reported between January 1990 and September 2004. Bold numbers relate to districts where data were evaluated during field surveys; italic numbers derive from incomplete reports to the Wildlife Division in Dar es Salaam. Dark green areas are National Parks; light green areas are Game Reserves. Asterisk marks Kilosa District. **b**, Number of attacks each year. The upward trend is significant ($P = 0.0030$), although data for 2004 are only complete to September. **c**, Principal contexts of attack on females (red) and males (blue), based on 218 attacks for which the context and victim's sex were known.

countries in the world. Mitigation of this fundamental conflict^{3,13} must be a priority for any lion conservation strategy in Africa.

Craig Packer*†, **Dennis Ikanda†**,
Bernard Kissui*†, **Hadas Kushnir***

*Department of Ecology, Evolution and Behavior,
University of Minnesota, St Paul, Minnesota
55108, USA

e-mail: packer@cbs.umn.edu

†Tanzania Wildlife Research Institute, Box 661,
Arusha, Tanzania

1. Kellert, S. R. *et al. Conserv. Biol.* **10**, 977–990 (1996).
2. Wilson, C. J. *Mammal Rev.* **34**, 211–232 (2004).
3. Ogada, M. O., Woodroffe, R., Oguge, N. O. & Frank, L. G. *Conserv. Biol.* **17**, 1521–1530 (2003).
4. Patterson, B. D., Kasiki, S. M., Selempo, E. & Kays, R. W. *Biol. Conserv.* **119**, 507–516 (2004).

5. Treves, A. *et al. Conserv. Biol.* **18**, 114–125 (2004).
6. Kerbis-Peterhans, J. C. & Gnoske, T. P. *J. East Afr. Nat. Hist.* **90**, 1–40 (2002).
7. Löe, J. & Röskopf, E. *Ambio* **33**, 283–288 (2004).
8. Chardonnet, P. *Conservation of the African Lion: Contribution to a Status Survey* (International Foundation for the Conservation of Wildlife, France, and Conservation Force, USA, 2002).
9. Bauer, H. & Van Der Merwe, S. *Oryx* **38**, 26–31 (2004).
10. Rushby, G. C. *No More the Tusker* (W. H. Allen, London, 1965).
11. Nicholson, B. *The Last of Old Africa* (Safari, Long Beach, California, 2001).
12. TRAFFIC. *Food For Thought: The Utilization of Wild Meat in Eastern and Southern Africa* (IUCN/WWF, Cambridge, UK, 1998).
13. Treves, A. & Karanth, K. U. *Conserv. Biol.* **17**, 1491–1499 (2003).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi: 10.1038/436791a

BROWNIAN MOTION

Absolute negative particle mobility

Noise effects in technological applications, far from being a nuisance, can be exploited with advantage — for example, unavoidable thermal fluctuations have found application in the transport and sorting of colloidal particles^{1–3} and biomolecules^{4–6}. Here we use a microfluidic system to demonstrate a paradoxical migration mechanism in which particles always move in a direction opposite to the net acting force ('absolute negative mobility') as a result of an interplay between thermal noise, a periodic and symmetric microstructure, and a biased alternating-current electric field. This counterintuitive phenomenon could be used for bioanalytical purposes, for example in the separation and fractionation of colloids, biological molecules and cells.

Newton's second law would seem to exclude absolute negative mobility as a response by a system at rest to a static force. However, such paradoxical behaviour has been experimentally observed in semiconductor devices⁷ and theoretically predicted in simplified stochastic model systems^{8,9}. Nonlinear dynamics are necessary to reconcile absolute negative mobility with Newton's law, and the system needs to operate far from equilibrium to avoid conflict with the second law of thermodynamics.

To provide a proof-of-principle for absolute negative mobility in a lab-on-a-chip, we designed a simple microfluidic device consisting of periodically arranged posts with alternating small and large gaps (Fig. 1; for methods, see supplementary information). The device is filled with negatively charged beads in an aqueous buffer. Electric fields are generated by applying a voltage along the *x*-axis, so that a positive voltage generates a positive force on the beads along the *x*-axis. Applying an alternating voltage, $U_{AC}(t)$, that switches periodically between $\pm U_0$, there is no net motion of the beads for symmetry rea-

sons. This set-up represents our unperturbed non-equilibrium system at rest. But what will be the average migration velocity v in the *x* direction in response to a static perturbation voltage, U_{DC} , superimposed on $U_{AC}(t)$?

To answer this question, we computed the electric field in the microstructure and, including thermal-noise effects, simulated the bead motion by stochastic differential equa-

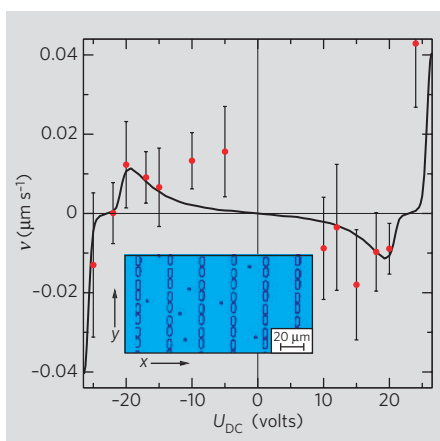


Figure 1 | Absolute negative mobility of polystyrene beads. Movement of 2-μm polystyrene beads in response to a static voltage, U_{DC} , superimposed to an alternating-current voltage that switches between ± 30 V every 25 s. Red dots: experimentally measured velocities, averaged over 40 beads and 200 s. Error bars are of statistical origin, but variabilities in the beads and microstructures also contribute. The line shows the theoretical response characteristics obtained from numerical simulations. Inset: partial view (optical micrograph image) of the poly(dimethylsiloxane) microstructure showing the rectangular posts and migrating beads (dark dots). In both the *x* and *y* directions, the gaps between the posts are alternately smaller and larger than the bead diameter. For movie, see supplementary information.

tions, adapted quantitatively to the microstructure geometry and the experimentally determined free diffusion and mobility of the microbeads (see supplementary information). The key feature of the resulting response curve (Fig. 1) is the negative slope that is symmetrical around the origin — a distinct and unambiguous signature of absolute negative mobility. We also measured the average bead velocity by tracking the beads using real-time video microscopy for different values of U_{DC} (see supplementary information) and found that the measured experimental response was in good agreement with theory (Fig. 1).

To explain how absolute negative mobility occurs, we consider the small gaps in the microstructure as 'traps' — electrical field lines can pass through them but the beads cannot. For $0 < U_{DC} < U_0$, the alternating total voltages $\pm U_0 + U_{DC}$ yield a back-and-forth motion of the beads along the *x* direction. Whenever a bead succeeds in passing through a large gap, it is trapped by the adjacent small gap in the *x* direction (Fig. 1, inset), unless it thermally diffuses sufficiently far in the *y* direction to proceed through another large gap. The smaller the voltage, the more time it has to do so and the farther it gets before being trapped. As $|+U_0 + U_{DC}| > |-U_0 + U_{DC}|$, the bead becomes trapped after moving forwards a short distance when $+U_0 + U_{DC} > 0$, whereas it travels backwards for a longer distance when $-U_0 + U_{DC} < 0$, resulting in absolute negative mobility.

We find that this response by a particle is very sensitive to particle size, even to the extent of allowing particles of different sizes to be steered in opposite directions (our unpublished results). Absolute negative mobility therefore holds promise for bioanalytical applications, for example in the sorting of cells, organelles and biochemical compounds.

Alexandra Ros*, **Ralf Eichhorn†**, **Jan Regtmeier***,
Thanh Tu Duong*, **Peter Reimann†**,
Dario Anselmetti*

*Experimental Biophysics & Applied Nanosciences, and †Condensed Matter Theory, Bielefeld University, 33501 Bielefeld, Germany
e-mail: alexandra.ros@physik.uni-bielefeld.de

1. Rousselet, J., Salome, L., Ajdari, A. & Prost, J. *Nature* **370**, 446–448 (1994).
2. Faucheux, L. P. & Libchaber, A. *Faraday Trans.* **91**, 3163–3166 (1995).
3. Marquet, C., Buguin, A., Talini, L. & Silberzan, P. *Phys. Rev. Lett.* **88**, 168301 (2002).
4. Han, J. & Craighead, H. G. *Science* **288**, 1026–1029 (2000).
5. Bader, J. S. *et al. Proc. Natl Acad. Sci. USA* **96**, 13165–13169 (1999).
6. Huang, L. R., Cox, E. C., Austin, R. H. & Sturm, J. C. *Analyt. Chem.* **75**, 6963–6967 (2003).
7. Keay, B. J. *et al. Phys. Rev. Lett.* **75**, 4102–4105 (1995).
8. Eichhorn, R., Reimann, P. & Hänggi, P. *Phys. Rev. Lett.* **88**, 190601 (2002).
9. Cleuren, B. & Van den Broeck, C. *Phys. Rev. E* **67**, 055101 (2003).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi: 10.1038/436792a

**Cover illustration**

The cells are from the recently developed robust *in vitro* hepatitis C virus (HCV) infection system. Pink staining indicates the presence of the HCV protein NS5A. (Courtesy of F. Chisari.) The crystal structure is of the HCV polymerase NS5B in complex with an inhibitor. (Courtesy of R. De Francesco & G. Migliaccio.)

Editor, Nature

Philip Campbell

Insights Publisher

Sarah Greaves

Insights Editor

Lesley Anson

Production Editor

Maria Hodges

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Layouts

Clifford Saunders

Sponsorship

Claire Hines

Production

Sue Gray

Marketing

Claire Aspinall

Editorial Assistant

Laura Shaw

HEPATITIS C

An estimated 3% of the world's population — more than 170 million people — are infected by the hepatitis C virus (HCV). Most infections become chronic: a condition that is incurable in many patients, leading to cirrhosis, end-stage liver disease and hepatocellular carcinoma. Current medical treatment options are limited, and 10,000 to 20,000 deaths a year in the United States are from hepatitis C. Indeed, chronic HCV infection is the most common cause of liver transplantation.

Despite the discovery of the virus by molecular biological methods more than 15 years ago, and the sequencing of its entire genome, our knowledge of the virus and the nature of the protective immune responses is limited. Researchers have been hampered by the lack of a robust cell-culture system yielding infectious virus until very recently, and the absence of a non-primate animal model.

Nevertheless, great strides have been made over the past few years, and much has been learned about the viral life cycle, immune response factors that confer protection, and the mechanisms by which the virus is able to evade the host immune response. New drugs are on the horizon and a protective vaccine may be within reach. This Insight brings together leading experts in hepatitis C to provide a snapshot of the field and discuss imminent developments.

We are pleased to acknowledge the financial support of Vertex Pharmaceuticals and Gilead Sciences, which together contributed towards the production of this Insight. As always, *Nature* carries sole responsibility for all editorial content and peer review.

Ursula Weiss, Senior Editor

INTRODUCTION

930 Unscrambling hepatitis C virus–host interactions

F. V. Chisari

REVIEW ARTICLES

933 Unravelling hepatitis C virus replication from genome to function

B. D. Lindenbach & C. M. Rice

939 Evasion of intracellular host defence by hepatitis C virus

M. Gale Jr & E. M. Foy

946 Adaptive immune responses in acute and chronic hepatitis C virus infection

D. G. Bowen & C. M. Walker

953 Challenges and successes in developing new therapies for hepatitis C

R. De Francesco & G. Migliaccio

961 Prospects for a vaccine against the hepatitis C virus

M. Houghton & S. Abrignani

967 Mechanism of action of interferon and ribavirin in treatment of hepatitis C

J. J. Feld & J. H. Hoofnagle

973 Hepatitis C and liver transplantation

R. S. Brown Jr

nature
insight

Unscrambling hepatitis C virus–host interactions

Francis V. Chisari¹

The human suffering exacted by the hepatitis C virus is enormous. Hundreds of thousands of people die each year from liver failure and cancer caused by this infection. There is no vaccine, and the available antiviral drugs are toxic, expensive and only partly effective. Progress has been hindered by the absence of cell culture and small-animal models of the infection. Nonetheless, recent advances have yielded several promising new antiviral drugs and enhanced the prospects of developing a vaccine. The recent development of a robust *in vitro* hepatitis C virus infection system will aid this search.

The hepatitis C virus (HCV) is a noncytopathic hepatotropic member of the Flaviviridae that causes acute and chronic hepatitis, and hepatocellular carcinoma (HCC; Fig. 1)¹. The liver is its primary target organ, and the hepatocyte is its primary target cell. More than 170 million people are currently infected with HCV². Acute infection is usually asymptomatic, making early diagnosis difficult. A notable feature of HCV infection is its tendency towards chronicity: ~70% of acute infections become persistent, and chronic cases are often associated with serious liver disease¹ (Fig. 2). As a result, HCV infection is a leading killer worldwide and the commonest cause of liver failure in the United States (see review in this issue by Brown, page 973).

In common with hepatitis B and human immunodeficiency (HIV) viruses, HCV is primarily transmitted percutaneously³. Before the development of diagnostic tests, the infection was commonly passed on through blood and related products⁴, haemodialysis⁵ and organ transplantation⁶. Today, HCV primarily affects injecting drug users and their sexual partners⁶. It is a particular problem in correctional facilities, where 20–40% of inmates are infected, in contrast to ~2% of the general population⁷. It is opportunistic in HIV-infected individuals, ~25% of whom are co-infected with HCV (this figure rises to 50–90% among injecting drug users)⁸. Co-infection causes higher HCV titres and a more rapid progression to cirrhosis⁸.

The virus and its life cycle

HCV infects only humans and chimpanzees; there are no small-animal models. Moreover, until recently, cell culture systems were not available. Most of our knowledge of HCV has been derived from surrogate experimental systems that approximate true infection and often preclude definitive interpretation. Nonetheless, much has been learned in the 16 years since the HCV genome was first cloned by Houghton and colleagues⁹.

The HCV genome is a 9.6-kilobase uncapped linear single-stranded RNA (ssRNA) molecule with positive polarity. It contains 5' and 3' untranslated regions (UTRs) including control elements required for translation and replication¹⁰ (see review in this issue by Lindenbach and Rice, page 933). The UTRs flank an uninterrupted open reading frame encoding a single polypeptide of 3,010 or 3,011 amino acids, which is processed into structural (C, E1, E2 and p7) and nonstructural (NS2, NS3, NS4A, NS4B, NS5A and NS5B) subunits by host and viral proteases^{11,12}.

The HCV life cycle is entirely cytoplasmic. Replication occurs through

a minus-strand intermediate in a membrane-bounded compartment¹³, yielding double-stranded RNA (dsRNA) intermediates. The replicative intermediates are fully exposed to the cell dsRNA-sensing machinery^{14,15} and induce strong innate cellular responses following infection.

Although much of our understanding of HCV replication is based on subgenomic and genomic replicon systems developed by Bartenschlager and colleagues¹⁶, little is known about the mechanisms and host functions involved in viral entry, uncoating, trafficking, assembly and egress. However, three independent groups have recently reported the development of a robust HCV infection system *in vitro*^{17–19}, which should make these processes experimentally accessible.

This new system is based on a unique HCV genome (JFH1) derived from the blood of a Japanese patient with fulminant hepatitis²⁰ and has extraordinary replicative capacity *in vitro*²¹ (see review in this issue by Lindenbach and Rice, page 933). The ability to perform reverse-genetics experiments will facilitate identification of the unique elements required for infectivity *in vitro* and help understand their specific roles. Moreover, by making each step of the viral life cycle experimentally accessible, this system will assist the development of antiviral drugs and help analyse the neutralizing and curative potential of candidate vaccines.

The host–virus relationship

In common with other persistent viruses, HCV does not kill the cells it infects, but triggers an immune-mediated inflammatory response (hepatitis) that either rapidly clears the infection or slowly destroys the liver, causing the development of HCC (see review in this issue by Bowen and Walker, page 946). The outcome is largely determined by the efficiency of the antiviral immune response. Host–virus interactions are ideally investigated in cell culture and small-animal models; the former are only now becoming available. Nonetheless, many of the factors that determine the outcome of HCV infection are beginning to come into focus (see reviews in this issue by Gale and Foy, and Bowen and Walker, pages 939 and 946, respectively).

Host–outcome determinants

Innate immune response

HCV spreads rapidly in the liver after inoculation^{22,23}, and so the innate immune response might be expected to influence the outcome of infection. Indeed, prospective genomic analysis of the intrahepatic

¹The Scripps Research Institute, La Jolla, California 92037, USA

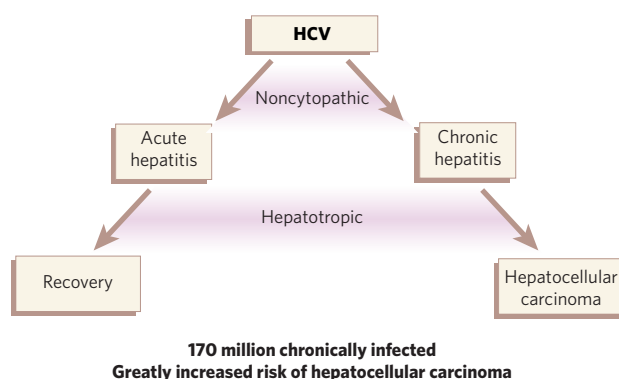


Figure 1 | Natural history of HCV infection. HCV is a noncytopathic virus that infects the liver and causes acute self-limited infection in 10–30% of patients with an associated inflammatory liver disease of variable severity that is mediated by the cellular immune response. In 70–90% of patients, however, HCV persists and causes chronic hepatitis with its life-threatening complications, including liver failure and hepatocellular carcinoma.

innate immune response in acutely infected chimpanzees suggests that HCV triggers a strong type-1 interferon (IFN- α/β) response as it spreads^{22,24}, but resists the effector functions of the downstream antiviral target genes that it induces. Importantly, the response is similar in animals that clear the infection and those that become persistently infected^{22,24}, implying that any influence on the outcome is indirect or obscure. Whatever its function, the innate intracellular immune response probably has a role in controlling HCV infection because the virus has developed several strategies to evade it (see review in this issue by Gale and Foy, page 939).

Adaptive immune response

The clearest determinants of the outcome of HCV infection are the magnitude, diversity and quality of the adaptive immune response. Viral clearance during self-limited infection is characterized by vigorous polyclonal CD4⁺ and CD8⁺ T-cell responses that are relatively weak and narrowly focused in chronically infected humans and chimpanzees. Moreover, the onset of viral clearance and liver disease coincide with that of the T-cell response and the entry of virus-specific T cells into the liver; primary failure to induce a T-cell response or functional exhaustion of an initially vigorous response predict viral persistence^{23,25}. However, the basis for variable immunological responsiveness to HCV has largely remained elusive. Indeed, we do not know whether the failure to respond vigorously in persistently infected subjects is caused by antigen overload during immunological priming, virus-induced defects in antigen presentation, hyperinduction of regulatory T cells, genetically determined restriction of the virus-specific T-cell repertoire or other causes (see review in this issue by Bowen and Walker, page 946). Therefore, whereas both primary and secondary immunological hyporesponsiveness to HCV seem to contribute to the establishment and maintenance of persistent infection, the reasons why they occur in selected subjects remain to be determined. Moreover, the virus can persist despite a multispecific CD4⁺ and CD8⁺ T-cell response^{23,25} by progressive mutational escape, which confirms the importance of the immune response in viral clearance and disease pathogenesis. An inbred mouse model of HCV infection would greatly facilitate these studies.

Viral outcome determinants

The viral factors influencing the outcome of HCV infection are beginning to come into focus.

HCV genotypes, replication and mutation rates

The six distinct genotypes of HCV show marked differences in geographic distribution, disease progression and response to therapy. However, the complex epidemiological differences in patient groups infected with each genotype make it difficult to ascribe variability in outcome to

the virus instead of the host (see review in this issue by Feld and Hoofnagle, page 967). The mutation rate of HCV is high (10^{-3} per nucleotide per generation), as is its replication rate ($\sim 10^{12}$ virions per day in humans)²⁶. This results in explosive expansion of the virus after inoculation and in the evolution of numerous viral quasispecies in each infected subject, which could influence the magnitude and efficacy of the antiviral immune response. Moreover, the virus produces a constant stream of escape variants that outrun the immune response and can eventually produce mutants with no corresponding receptors in the immunological repertoire^{23,25}. The influence of these parameters on the outcome of infection has been studied in a few acutely infected humans and chimpanzees and in many chronically infected individuals. The results show that B- and T-cell escape mutants are selected by the immune response during HCV infection and probably contribute to viral persistence (see review in this issue by Bowen and Walker, page 946).

Viral evasion strategies

The primary immune-evasion strategies fall into two distinct categories: subversion of the IFN response induced by the virus and mutational escape from the adaptive immune response.

According to the first strategy, when HCV infects the liver, it triggers the production of IFN and a range of antiviral genes that should control the infection — but do not^{22,27}. In fact, HCV seems to be resistant to these antiviral pathways, at least in the HCV replicon system, and several structural and nonstructural proteins have been shown to inhibit nonoverlapping functions of the innate immune response (see review in this issue by Gale and Foy, page 939). For example, the Gale group has shown that NS3/4A can block the phosphorylation and effector action of IFN regulatory factor 3 (IRF3)²⁸ by inactivating signalling by retinoic-acid-inducible gene I (RIG-I)²⁹, a cytoplasmic dsRNA-binding protein that activates cellular kinases that stimulate IRF3 (ref. 15). So, HCV seems to use several strategies to actively evade the immune responses that it induces. However, these evasion strategies were defined either biochemically or in transfected cell-culture systems, not in infected cells. It is premature to assume that they occur during natural infection until they have been validated *in vivo* or at least in the newly developed tissue-culture model of HCV infection.

Regarding the second strategy, mutational inactivation of B- and T-cell epitopes is common in HCV infection (see review in this issue by Bowen and Walker, page 946). B-cell epitopes are concentrated in the hypervariable region of the E2 protein³⁰, probably allowing the virus to persist in the presence of antibody that is neutralizing for its ancestors. The T-cell epitope mutations span the viral polyprotein³¹, often in residues that bind to major histocompatibility complex (MHC) molecules or are otherwise involved in antigen presentation. Mutations also occur in residues engaged by the T-cell receptor (TCR), making infected cells invisible to T cells expressing the corresponding TCR³². Although mutational escape probably contributes to the persistence of the virus, it is less clear whether it determines the outcome. Several groups have shown an association between certain human leukocyte antigen (HLA) alleles and the outcome of HCV infection²⁵. These differences might influence the breadth of the TCR repertoire and the ease with which the virus can escape. Confirmation of this hypothesis would be facilitated by an inbred mouse model of HCV infection.

What about treatment?

The standard treatment for chronic HCV infection is pegylated IFN- α plus ribavirin³³ (see review in this issue by Feld and Hoofnagle, page 967). Although the mechanism of action of these drugs is debated, with both antiviral and immunostimulatory mechanisms being implicated³⁴, the sustained response (cure) rates are far from ideal. Moreover, there is substantial associated toxicity³⁵, and the likelihood of success depends on viral and host factors that are often beyond the control of patients and physicians. Clearly, more effective and less toxic treatment regimens are needed.

Thanks to the HCV replicon system, much recent effort has been directed towards developing drugs that inhibit viral replication. Sev-

eral promising small-molecule inhibitors of the NS3/4A protease and the NS5B polymerase are in development (see review in this issue by De Francesco and Migliaccio, page 953). Early testing has uncovered strong antiviral activity both *in vitro* and in patients³⁶. As expected, escape variants have been rapidly selected by each drug, indicating that drug cocktails will probably be required to control this infection.

The hope is to target all aspects of the viral life cycle therapeutically, including those that are inaccessible using the replicon system. Indeed, this is becoming possible owing to the *in vitro* HCV infection model mentioned above^{17–19}.

De Francesco and Migliaccio also describe efforts to stimulate the innate immune response in chronically infected subjects. Furthermore, several groups are attempting to activate the adaptive immune response with therapeutic vaccines³⁷ (see reviews in this issue by Houghton and Abrignani, and Bowen and Walker, pages 961 and 946, respectively). Harnessing the immune response makes sense, as it is curative in subjects who spontaneously clear HCV infection and could potentially eliminate the drug-resistant viral clones selected by emerging antiviral drugs.

Prospects for a preventive vaccine

Developing a preventive vaccine is arguably the ultimate objective of this research. Traditional vaccine development based on the antibody response preventing incoming viruses from reaching their cellular targets has been challenging for mutation-prone HCV, which generates numerous antibody escape mutants³⁸.

Nonetheless, a vaccine might be within reach. Chimpanzees that are rechallenged after clearance of a primary HCV infection are protected against homologous and heterologous viral isolates³⁹. Moreover, those immunized with an adjuvanted recombinant HCV envelope vaccine that elicits E2 glycoprotein-specific antibodies and T-cell help are largely protected from chronic infection when challenged with a heterologous viral inoculum (see review in this issue by Houghton and Abrignani, page 961). This is a considerable advance, even though the immunized chimpanzees become acutely infected, because the morbidity of HCV is largely a consequence of chronic infection, from which they are protected.

As spontaneous viral clearance during acute HCV infection is characterized by a vigorous broadly reactive CD4⁺ and CD8⁺ T-cell response, much effort is being made to develop T-cell-based vaccines. Approaches include DNA-prime–protein boost, DNA-prime–poxvirus boost, recombinant adenovirus and Semliki Forest virus vectors, and recombinant α -viral particles (see review in this issue by Houghton and Abrignani, page 961). Despite their limitations, these techniques potentially offer enhanced immunogenicity, the ability to induce antibody as well as T-cell responses and the flexibility to stimulate cross-protective immunity.

Where do we go from here?

We arrive at the end of this article with more questions than answers as we try to understand the viral life cycle, the factors that determine the outcome of infection and how to develop a vaccine and more effective drugs for the prevention and treatment of HCV infection. We know that the virus infects the liver, but how many hepatocytes are infected, are other cells affected and does it disrupt luxury functions or have direct cytopathic effects? We know that HCV spreads and mutates rapidly and that a huge number of virions are produced every day. But we don't know the extent to which small differences in these parameters allow it to outrun the immune response and persist in certain patients. The virus survives despite a strong innate intracellular IFN response and has evolved the means to evade it, yet is the outcome determined by an undefined counter-regulatory gene or isoform expressed only in subjects who clear the infection? We know that certain HCV proteins can blunt the antiviral effects of IFN in transfected cells that overexpress them, but does this occur during natural infection *in vivo*? Viral clearance is associated with a sustained vigorous polyclonal adaptive immune response, yet why do some individuals mount such a response and others do not? Such questions must be answered to understand the outcome determinants, and prevent and terminate persistent HCV infection. New experimental systems are urgently needed if we wish to answer those difficult ques-

tions. The recent development of a cell-culture system supporting robust HCV infection^{17–19} is a huge step in the right direction. ■

1. Hoofnagle, J. H. Course and outcome of hepatitis C. *Hepatology* **36**, S21–S29 (2002).
2. Alter, H. J. & Seeff, L. B. Recovery, persistence, and sequelae in hepatitis C virus infection: a perspective on long-term outcome. *Semin. Liver Dis.* **20**, 17–35 (2000).
3. Alter, M. J. Epidemiology of hepatitis C. *Eur. J. Gastroenterol. Hepatol.* **8**, 319–323 (1996).
4. Alter, M. J. Epidemiology of hepatitis C. *Hepatology* **26**, 62S–65S (1997).
5. Alter, M. J. Hepatitis C virus infection in the United States. *J. Hepatol.* **31** (Suppl. 1), 88–91 (1999).
6. Alter, M. J. Prevention of spread of hepatitis C. *Hepatology* **36**, S93–S98 (2002).
7. Spaulding, A., Greene, C., Davidson, K., Schneidermann, M. & Rich, J. Hepatitis C in state correctional facilities. *Prev. Med.* **28**, 92–100 (1999).
8. Sulkowski, M. S., Mast, E. E., Seeff, L. B. & Thomas, D. L. Hepatitis C virus infection as an opportunistic disease in persons infected with human immunodeficiency virus. *Clin. Infect. Dis.* **30** (Suppl. 1), S77–S84 (2000).
9. Choo, Q. L. et al. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**, 359–362 (1989).
10. Sarnow, P. Viral internal ribosome entry site elements: novel ribosome-RNA complexes and roles in viral pathogenesis. *J. Virol.* **77**, 2801–2806 (2003).
11. Lohmann, V., Koch, J. O. & Bartenschlager, R. Processing pathways of the hepatitis C virus proteins. *J. Hepatol.* **24**, 11–19 (1996).
12. Penin, F., Dubuisson, J., Rey, F. A., Moradpour, D. & Pawlotsky, J. M. Structural biology of hepatitis C virus. *Hepatology* **39**, 5–19 (2004).
13. Moradpour, D. et al. Membrane association of the RNA-dependent RNA polymerase is essential for hepatitis C virus RNA replication. *J. Virol.* **78**, 13278–13284 (2004).
14. Samuel, C. E. Antiviral actions of interferons. *Clin. Microbiol. Rev.* **14**, 778–809 (2001).
15. Yoneyama, M. et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
16. Bartenschlager, R. Hepatitis C virus replicons: potential role for drug development. *Nature Rev. Drug Discov.* **1**, 911–916 (2002).
17. Zhong, J. et al. Robust hepatitis C virus infection *in vitro*. *Proc. Natl Acad. Sci. USA* **102**, 9294–9299 (2005).
18. Lindenbach, B. D. et al. Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623–626 (2005).
19. Wakita, T. et al. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Med.* **11**, 791–796 (2005).
20. Kato, T. et al. Sequence analysis of hepatitis C virus isolated from a fulminant hepatitis patient. *J. Med. Virol.* **64**, 334–339 (2001).
21. Kato, T. et al. Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology* **125**, 1808–1817 (2003).
22. Wieland, S. F. & Chisari, F. V. Stealth and cunning: hepatitis B and hepatitis C. *J. Virol.* **79**, 9369–9380 (2005).
23. Rehermann, B. & Nascimbeni, M. Immunology of hepatitis B virus and hepatitis C virus infection. *Nature Rev. Immunol.* **5**, 215–229 (2005).
24. Bigger, C. B. et al. Intrahepatic gene expression during chronic hepatitis C virus infection in chimpanzees. *J. Virol.* **78**, 13779–13792 (2004).
25. Shoukry, N. H., Cawthon, A. G. & Walker, C. M. Cell-mediated immunity and the outcome of hepatitis C virus infection. *Annu. Rev. Microbiol.* **58**, 391–424 (2004).
26. Neumann, A. U. et al. Hepatitis C viral dynamics *in vivo* and the antiviral efficacy of interferon- α therapy. *Science* **282**, 103–107 (1998).
27. Lanford, R. E. & Bigger, C. Advances in model systems for hepatitis C virus research. *Virology* **293**, 1–9 (2002).
28. Foy, E. et al. Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. *Science* **300**, 1145–1148 (2003).
29. Foy, E. et al. Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-1 signaling. *Proc. Natl Acad. Sci. USA* **102**, 2986–2991 (2005).
30. Mondelli, M. U. et al. Hypervariable region 1 of hepatitis C virus: immunological decoy or biologically relevant domain? *Antiviral Res.* **52**, 153–159 (2001).
31. Bowen, D. G. & Walker, C. M. Mutational escape from CD8⁺ T cell immunity: HCV evolution, from chimpanzees to man. *J. Exp. Med.* **201**, 1709–1714 (2005).
32. Meyer-Olson, D. et al. Limited T cell receptor diversity of HCV-specific T cell responses is associated with CTL escape. *J. Exp. Med.* **200**, 307–319 (2004).
33. McHutchison, J. G. & Fried, M. W. Current therapy for hepatitis C: pegylated interferon and ribavirin. *Clin. Liver Dis.* **7**, 149–161 (2003).
34. Pawlotsky, J. M. Mechanisms of antiviral treatment efficacy and failure in chronic hepatitis C. *Antiviral Res.* **59**, 1–11 (2003).
35. Strader, D. B., Wright, T., Thomas, D. L. & Seeff, L. B. Diagnosis, management, and treatment of hepatitis C. *Hepatology* **39**, 1147–1171 (2004).
36. Lamarre, D. et al. An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **426**, 186–189 (2003).
37. Nevens, F. et al. A pilot study of therapeutic vaccination with envelope protein E1 in 35 patients with chronic hepatitis C. *Hepatology* **38**, 1289–1296 (2003).
38. Shimizu, Y. K. et al. Neutralizing antibodies against hepatitis C virus and the emergence of neutralization escape mutant viruses. *J. Virol.* **68**, 1494–1500 (1994).
39. Bassett, S. E. et al. Protective immune response to hepatitis C virus in chimpanzees rechallenged following clearance of primary infection. *Hepatology* **33**, 1479–1487 (2001).

Acknowledgments The author thanks S. Wieland for critical reading of the manuscript and help with manuscript preparation. The work from the laboratory of the author that is cited in this paper was supported by US National Institutes of Health grants.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares that he has no competing financial interests. Correspondence should be addressed to fchisari@scripps.edu.

Unravelling hepatitis C virus replication from genome to function

Brett D. Lindenbach¹ and Charles M. Rice¹

Since the discovery of the hepatitis C virus over 15 years ago, scientists have raced to develop diagnostics, study the virus and find new therapies. Yet virtually every attempt to dissect this pathogen has met with roadblocks that impeded progress. Its replication was restricted to humans or experimentally infected chimpanzees, and efficient growth of the virus in cell culture failed until very recently. Nevertheless hard-fought progress has been made and the first wave of antiviral drugs is entering clinical trials.

In the mid-1970s, it was noticed that the world's supply of blood was contaminated with an unidentified agent causing post-transfusion non-A, non-B hepatitis¹. Yet it was not until 1989 that the first sequences of hepatitis C virus (HCV) were reported². The difficulty in identifying the virus and in detecting viral RNA and antigens in infected tissues left the impression that HCV replicated poorly *in vivo*. This is not so. Chronically infected patients have viral loads that typically range from 10^3 – 10^7 genomes per ml of serum. Mathematical modelling of viral dynamics during treatment with interferon- α (IFN- α) indicates that HCV virions turn over rapidly (with a half-life about 3 h), and up to about 10^{12} viruses are produced per day in an infected person³. This is about 100-fold greater than the rate reported for HIV.

High viral loads are observed during the first few weeks of HCV infection, but inflammatory processes leading to liver injury are delayed, usually occurring after 2–3 months⁴. Liver transplant recipients generally have favourable short-term outcomes despite efficient allograft reinfection and high levels of viraemia owing to immunosuppression. These observations have led to the idea that HCV is relatively noncytotoxic and that liver disease is immune-mediated. Although the liver is the major site of HCV replication, evidence exists for extrahepatic reservoirs including peripheral blood lymphocytes (reviewed in ref. 5), epithelial cells in the gut⁶ and the central nervous system⁷. With the hepatitis B virus immunohistochemistry can be used to identify infected hepatocytes reliably, but we do not have a clear picture of the number of HCV-infected hepatocytes in the liver or the characteristics of an HCV-infected cell. Nevertheless, gene-profiling studies of HCV-infected livers indicate that this organ is a veritable battleground of ongoing viral replication and host antiviral defences^{8–11}. Although the origin of HCV and the timing of its introduction into the human population are not known, the high error rate of RNA-dependent RNA replication and the battle between virus and host have generated remarkable global diversity. HCV is currently divided into six major genotypes with numerous subtypes and exists as a quasispecies swarm within the infected individual¹².

We will use an idealized HCV life cycle (Fig. 1) as a framework for discussing the current state of our knowledge. Enveloped virus particles interact with specific surface receptors and are probably internalized. Fusion of the viral and cellular membranes, presumably triggered by the low pH of the endocytic compartment, leads to the release of a single-stranded (ss), positive-sense RNA genome into the cytoplasm of a newly infected cell. This genome serves multiple roles within the

virus life cycle: first, as a messenger RNA (mRNA) for translation of the viral proteins; second as a template for RNA replication; and third, as a nascent genome packaged within new virus particles. Virions presumably form by budding into the endoplasmic reticulum (ER) and leave the cell through the secretory pathway.

Researchers have followed each aspect of the virus life cycle in turn. Just as infection starts from an HCV genome entering the cytoplasm and progressing through translation, replication and particle production, our understanding has progressed from having a genome sequence to understanding translation and the viral gene products, characterizing RNA replication, and establishing systems to characterize virus particles and infectivity. In this review, we summarize the current understanding of HCV replication with special emphasis on recent developments. As space is limited, we cannot be comprehensive; readers may wish to consult other reviews for detail and breadth^{5,13,14}.

Initial studies: HCV translation and polyprotein processing

The identification of HCV yielded a partial viral genome sequence⁵. Research in the early 1990s focused on dissecting HCV gene expression and characterizing the gene products. Much has been learned about the biochemistry of three key enzymes, and structural information is now available at the atomic level for roughly half of the protein-coding region.

Translation of the HCV genome, which lacks a 5' cap, depends on an internal ribosome entry site (IRES) within the 5' noncoding region (NCR). The HCV IRES binds 40S ribosomal subunits directly and avidly, bypassing the need for pre-initiation factors, and inducing an mRNA-bound conformation in the 40S subunit¹⁵. The IRES–40S complex then recruits eukaryotic initiation factor (eIF) 3 and the ternary complex of Met-tRNA–eIF2–GTP to form a non-canonical 48S intermediate, before a kinetically slow transition to the translationally active 80S complex^{16,17}.

Once initiated, translation of the HCV genome produces a large polyprotein that is proteolytically cleaved to produce 10 viral proteins (Fig. 2a). The amino-terminal one-third of the polyprotein encodes the virion structural proteins: the highly basic core (C) protein, and glycoproteins E1 and E2. After the structural region comes a small integral membrane protein, p7, which seems to function as an ion channel^{18,19}. The remainder of the genome encodes the nonstructural (NS) proteins NS2, NS3, NS4A, NS4B, NS5A and NS5B, which coordinate the intracellular processes of the virus life cycle. The structural

¹Center for the Study of Hepatitis C, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

proteins mature by signal peptidase cleavages between C/E1, E1/E2 and E2/p7. In addition, signal-peptide peptidase releases core from the E1 signal peptide. Within the NS region, the p7/NS2 junction is also cleaved by signal peptidase. Further proteolytic processing within the NS region occurs through the action of two viral enzymes, the NS2 autoprotease, which cleaves at the NS2/3 junction; and the NS3-4A serine protease, which cleaves at all downstream sites (Fig. 2a, b). HCV also encodes a small protein, called F (frame shift) or ARFP (alternative reading frame protein), that can be produced by ribosomal frame shifting into an alternative reading frame within the core gene (reviewed in ref. 20).

HCV encodes two remarkable proteases

The carboxy-terminal two-thirds of NS2 contain the catalytic triad of a cysteine protease. NS2/3 cleavage requires these residues as well as the downstream expression of the NS3 serine protease domain, although NS3-4A protease activity is dispensable for NS2/3 processing. The requirement for NS3 may have to do with correct folding, as NS2/3 cleavage is enhanced by Zn^{2+} , which has a structural role in stabilizing the NS3 fold.

NS3 is a multifunctional protein, with an N-terminal serine protease domain and a C-terminal RNA helicase/NTPase domain. Both enzyme activities have been well characterized, and high-resolution structures have been solved¹⁴. The serine protease domain has a typical chymotrypsin-like fold, positioning three catalytically active residues at the surface interface between two β -barrel domains (Fig. 3a). For complete folding and serine protease activity NS3 requires the intercalation of a β -strand present in NS4A (refs 21–23). NS4A is a small (54-amino-acid) protein that anchors NS3 to cellular membranes through an N-terminal hydrophobic peptide²⁴. Proper folding of the serine protease domain also requires coordination of Zn^{2+} by three cysteine residues distal from the active site. NS3 has an unusually shallow substrate-binding pocket for a serine protease, which has challenged efforts to obtain specific inhibitors. Nevertheless, the discovery of product inhibition led to the development of potent inhibitors that block NS3 serine protease activity, NS protein processing, and HCV RNA replication. For further details, see the accompanying article by De Francesco and Migliaccio (page 953).

Dissecting the structure and function of HCV NS proteins

The C terminus of NS3 encodes a DEXH/D-box RNA helicase. These enzymes use the energy of NTP hydrolysis to unwind double-stranded RNA, and NS3 unwinds RNA and DNA homoduplexes and heteroduplexes in a 3' to 5' direction²⁵. The helicase mechanism is not yet fully understood, but recent kinetic analyses show that the NS3 helicase behaves like a ratcheting two-stroke motor²⁶ and seems to function as a dimer that incrementally rips apart 18-base-pair stretches of substrate RNA²⁷. In addition, NS3 helicase activity can be regulated by interactions between the serine protease and helicase domains of NS3 (refs 28, 29), indicating that these two enzyme activities may be somehow coordinated during replication. The function of the HCV helicase is not known, but it may be involved in the initiation of RNA synthesis on the HCV genome RNA, which contains stable 3'-terminal secondary structure, in dissociation of nascent RNA strands from their template during RNA synthesis, or in displacement of proteins or other *trans*-acting factors from the RNA genome.

NS5A is a fascinating protein. Many cellular proteins interact with NS5A, although their functional relevance is largely unclear^{30–32}. NS5A is phosphorylated on multiple serine residues by cellular kinases and can be found in hypophosphorylated (56 kDa) and hyperphosphorylated (58 kDa) forms. Major phosphorylation sites have been determined for a few HCV isolates^{33,34}, and kinases capable of phosphorylating NS5A have been identified. These include AKT, p70S6K, MEK1, MKK6, cAMP-dependent protein kinase A- α and casein kinase II^{35–38}. It is not yet clear which kinases are involved in generating the different phosphoforms of NS5A, nor which phosphorylation sites are functionally relevant, and the role of NS5A phos-

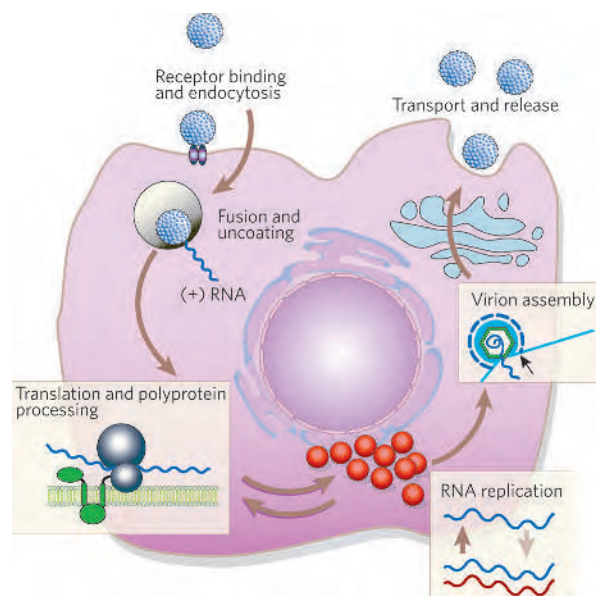


Figure 1 | HCV life cycle. After entry into the cell and uncoating, the HCV genome functions in three main roles: translation, replication and packaging into nascent virions.

phorylation remains an area of intense interest.

NS5A associates with membranes through an N-terminal amphipathic α -helix³⁹ and contains three distinct structural domains⁴⁰. Biochemical and structural analyses revealed that domain I (residues 1–213) forms a dimer with a novel fold and coordinates Zn^{2+} through a unique motif^{40,41}. This architecture reveals conserved external surfaces that might interact with other proteins, as well as a highly basic channel that might be involved in RNA binding (Fig. 3b). In addition, a disulphide bond was discovered between cysteine residues 142 and 190, which is a rare modification for cytosolic proteins and could represent another form of regulation. Unfortunately, structural information is not yet available for domains II and III, which contain determinants influencing the efficiency of RNA replication (see below) and NS5A phosphorylation. The C-terminal domain is not well conserved and seems quite flexible^{42,43}.

The workhorse of the HCV RNA replication machinery is NS5B, which encodes the RNA-dependent RNA polymerase (RdRP). Primer-dependent and *de novo* (unprimed) initiation of RNA synthesis have been demonstrated for this protein. On the basis of the replication strategy of other positive-strand RNA viruses, HCV RNA replication probably involves *de novo* initiation by a multiprotein complex (replicase). NS5B has a typical 'right hand' polymerase structure, with catalytic sites in the base of the palm domain, surrounded by thumb and finger domains¹⁴. These latter domains fully encircle the active site, creating a channel for binding to a ssRNA template (Fig. 3c). In addition, a β -hairpin structure protrudes from the thumb toward the active site and is likely to be involved in correct positioning of the template⁴⁴. The overall structure of NS5B is remarkably similar to the RdRP of bacteriophage $\phi 6$ (ref. 45), and cocrystallization of these enzymes with model substrates and nucleoside triphosphates has yielded a credible model for *de novo* initiation (reviewed in ref. 46). NS5B also has a low-affinity GTP-binding site, distal from the active site, which is thought to be an allosteric regulator of the finger-thumb interaction⁴⁷. NS5B is tethered to membranes by a C-terminal peptide anchor⁴⁸ and interacts with itself to form higher-order RdRP complexes that may have functional relevance to the membrane-bound replicase, described below⁴⁹.

The replication era: reverse genetic systems for HCV

Once the highly conserved 3' terminus of HCV was discovered and the genome sequence was completed^{50,51}, reverse genetics with HCV became possible. Functional complementary DNA clones were assem-

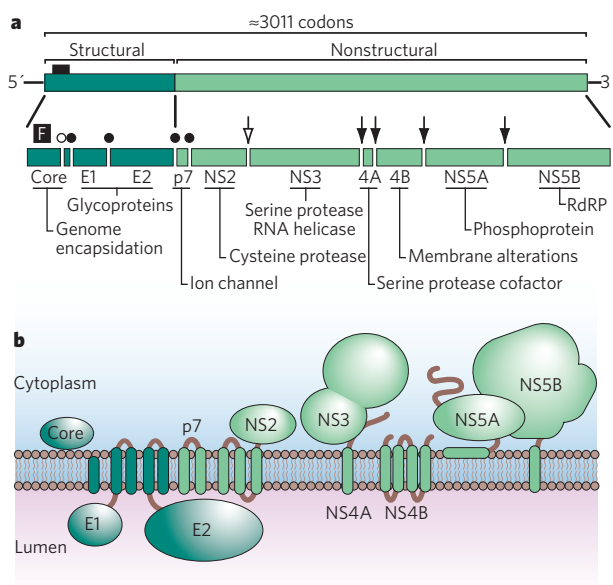


Figure 2 | HCV genes and gene products. **a**, The structure of the viral genome, including the long open reading frame encoding structural and nonstructural genes, and 5' and 3' NCRs. The polyprotein processing scheme is shown below. Closed circles refer to signal peptidase cleavage sites; the open circle refers to the signal peptide peptidase cleavage site. All other terms are defined in the text. **b**, The topology of HCV proteins with respect to a cellular membrane.

bled and shown to be infectious by direct intrahepatic injection of RNA transcripts into chimpanzees^{52,53}. These infectious clones were used to show that all viral enzyme activities, the p7 gene and the correct genomic 3' end are necessary for HCV replication *in vivo*^{54–56}. In contrast, the hypervariable N-terminal region of E2 is dispensable⁵⁷. Essentially clonal infections could be initiated from transcribed RNA, providing a well-defined genetic starting point to study virus evolution and immune responses to infection⁵⁸. With functional cDNA clones in hand and the ability to make unlimited quantities of infectious HCV RNA, much effort was devoted to searching for permissive HCV cell culture conditions. However, despite their great utility for *in vivo* studies, these initial chimpanzee infectious transcripts failed to replicate in cell culture.

A breakthrough for the field came in 1999 when Lohmann *et al.* reported selection of the first functional 'subgenomic' replicons in cell culture⁵⁹. These replicons consisted of a genotype 1b HCV RNA engineered to express a selectable marker gene, *Neo*, in place of the structural protein coding region, which was not expected to be required for RNA replication. To direct expression of NS proteins, a heterologous viral IRES was inserted after the neomycin resistance cassette (Fig. 4). After RNA transfection into a human hepatoma line, Huh-7, a few drug-selected colonies grew out that contained replicating HCV RNAs. The replicon system provided an important tool for studying HCV RNA replication and established a functional cell-based system for evaluating potential antiviral compounds.

The next discovery was that replicon RNAs harboured culture-adaptive mutations, often in NS5A, that increased RNA replication and Neo-transduction efficiency by up to 10,000-fold⁶⁰. Adaptive changes were subsequently mapped throughout the NS region, including NS3, NS4B, NS5A and NS5B (reviewed in ref. 13). In addition to the original genotype 1b isolate, called Con1, replicons have now been established for other 1b isolates and for genotypes 1a and 2a, and HCV RNA replication has been achieved in Hela, 293, HepG2 and even mouse hepatoma cell lines¹³. Thus, whereas it was once thought that HCV replication might be restricted to the environment provided by hepatocytes, it is now clear that a number of cells can support RNA replication. As discussed by Gale and Foy (see page 939, in this issue) the interplay between HCV and innate cellular

antiviral pathways may be a major intracellular determinant of HCV tropism.

Mechanisms of HCV RNA replication

Characterization of cell-culture-adaptive mutations is an area of great interest, as they are likely to teach us about the interface between HCV replication and the host cellular environment. Adaptive mutations in NS4B, NS5A or NS5B strongly enhance replication but are incompatible with each other, whereas adaptive changes in NS3 tend to be weak and cooperatively enhance replication when combined with strongly adaptive mutations¹³. Furthermore, adaptive mutations in NS3 and NS5B map to surface residues distant from the enzyme active sites, suggesting that these changes are likely to affect interactions between NS proteins and/or cellular factors. A striking feature of highly adaptive mutations is their tendency to decrease NS5A hyperphosphorylation, and many adaptive changes map to NS5A residues implicated in this process^{60–62}. Similarly, increased hyperphosphorylation of NS5A correlates with lower replication levels^{42,63,64} and decreased interaction between NS5A and a cellular factor, human vesicle-associated membrane-protein associated protein A (hVAP-A)⁶³. This vesicle-sorting protein has also been shown to localize HCV NS proteins in cholesterol-rich, detergent-resistant membranes that may be subcellular microenvironments for HCV RNA replication⁶⁵.

As for all positive-strand RNA viruses, HCV RNA replication occurs in association with altered cytoplasmic membranes. In replicon-bearing cells, HCV NS proteins, RNA and RdRP activity associate with ultrastructural vesicular structures termed the 'membranous web', which also resembles the membrane alterations seen in hepatocytes from HCV-infected liver^{66,67}. The integral membrane protein NS4B is sufficient to induce membranous web formation and has been proposed to serve as a scaffold for replication complex assembly⁶⁶. As mentioned, several adaptive mutations have been mapped to this protein, and NS4B has been found to encode a GTPase activity that may be related to its membrane-altering ability⁶⁸. Association of the HCV replicase with the membranous web can be followed in live cells by using subgenomic replicons with green fluorescent protein inserted in the NS5A domain III⁴³.

Recent experiments have shed new light on the interplay between cellular membranes, lipid metabolism and HCV replication. In cell culture, HCV RNA replication is stimulated by increasing the availability of saturated and monounsaturated fatty acids, and inhibited by polyunsaturated fatty acids or inhibitors of fatty acid synthesis⁶⁹. These results suggest that membrane fluidity is important for the function of the membranous web. In addition, two groups have shown that inhibition of protein geranylgeranylation leads to the disassembly of HCV replication complexes and strong inhibition of HCV RNA replication^{69,70}. In the absence of appropriate prenylation motifs in HCV proteins, these results suggest that a geranylgeranylated cellular protein participates in HCV replication and could be amenable to pharmacological manipulation. Recently, the geranylgeranylated host protein FBL-2 was shown to interact with NS5A and to be required for HCV RNA replication⁷¹. It is intriguing that FBL-2, which contains an 'F-box' motif, is likely to be involved in targeting proteins for degradation, although the identity of a relevant substrate is currently unknown.

An exciting area of progress has been the delineation of *cis*-acting RNA elements that guide viral replication. Nearly the entire 5' NCR is needed for efficient RNA amplification, although a minimal replication element exists within the first 120 nucleotides (refs 72–75). As this overlaps with the IRES, there is considerable interest in understanding how this region might modulate translation and replication, which are unlikely to occur simultaneously on the same RNA template. The 3' NCR has been found to contain a nonessential variable region, a poly-U/UC tract that must be more than 26 nucleotides long, followed by a highly conserved and essential 3' X domain^{55,57,76}. Recently a conserved stem-loop structure within the NS5B coding region, 5BSL3.2, was found to be required for RNA replication⁷⁸. Further studies indicated that 5BSL3.2 forms functionally important long-distance base pairs

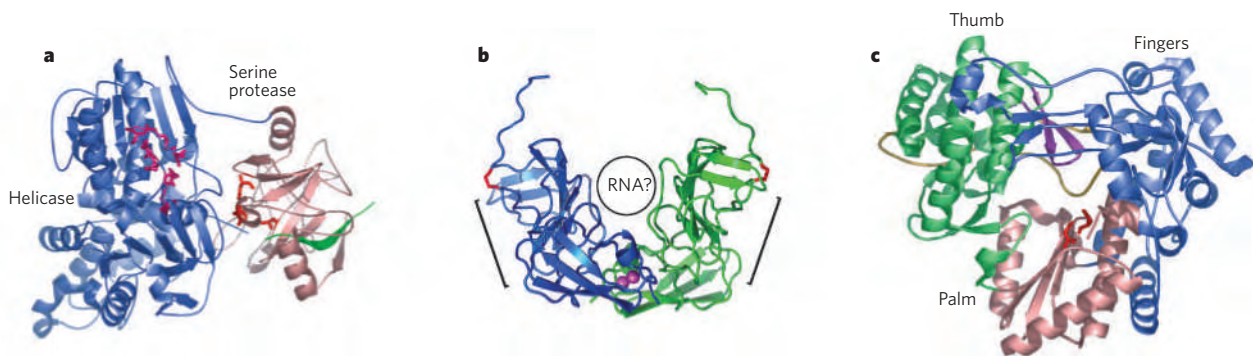


Figure 3 | HCV NS protein structures. **a**, NS3. The serine protease, NS4A cofactor and RNA helicase domains are shown in pink, green and blue, respectively. The serine protease and RNA helicase active site residues are indicated in red. **b**, NS5A domain I. Shown is a dimer, as seen in the crystal structure. Individual subunits are shown in blue and green, with their C termini (that is, leading into domain II) pointing upwards. The N termini, which presumably face the membrane, are at the bottom. The purple spheres represent Zn^{2+} ions. Disulphide bonds are indicated in

red. Brackets indicate highly conserved surfaces. A basic groove, which may bind RNA, is also indicated. **c**, NS5B. Shown is the typical 'right hand' model of the RdRP, with palm, fingers and thumb domains in pink, blue and green, respectively. The C-terminal region, which is not part of the RdRP, is shown in yellow. Note the extensive interactions between the finger and thumb domains. In addition, a β -hairpin is shown in purple, and active site residues Asp 220 and Asp 318 are shown in red.

with the 3'X domain⁷⁹. The *trans*-acting factors and the replication steps requiring this 'kissing' interaction remain to be determined.

Little is known about the process of HCV RNA synthesis within the replication complex. By inference from related viruses, RNA synthesis is likely to be semiconservative and asymmetric: the positive-strand genome serves as a template to make a negative-strand intermediate; the negative strand then serves as a template to produce multiple nascent genomes. RdRP activity can be detected in extracts from replicon-bearing cells, although this seems to reflect elongation of co-extracted templates rather than *de novo* initiation. Nascent products from these reactions are protected from nuclease digestion by a detergent- and protease-sensitive factor^{80,81}. Interestingly, protease treatment of permeabilized cells destroyed most NS proteins without compromising RdRP activity, suggesting that only a small fraction of NS proteins is actively engaged in RNA replication⁸⁰. The function of the 'excess' NS proteins, the composition of the HCV replicase and the possible reasons for physically sequestering the replicase (such as evasion of dsRNA sensors or antiviral effectors like RNAi) remain intriguing areas for future study.

Completing the virus life cycle: extracellular virions

Important questions about the nature of the infectious virus particle, the pathway of virus entry, and the assembly of viral structural proteins and RNA into new virus particles are still largely unanswered. Fortunately, new technologies have emerged that extend our reach into these formerly intractable areas.

HCV particles present in clinical samples have been partly characterized. Enveloped virions are sensitive to detergent and to chloroform, with a diameter of about 50 nm. HCV RNA and infectivity in chimpanzees have been followed in isopycnic density gradients. HCV exhibits unusual heterogeneity in buoyant density, with the peak of infectivity near 1.10 g ml^{-1} . This is surprisingly low even for an enveloped virus. This heterogeneity and low density have been explained in part by the association of HCV particles with serum components such as immunoglobulins and β -lipoproteins^{5,13}.

Expression and processing of the HCV structural gene products were characterized in early heterologous expression studies. Core protein was localized to the cytoplasmic surface of the ER and lipid droplets, and occasionally the cell nucleus^{82–84}. A central hydrophobic domain is responsible for the membrane association of the core, whereas the high pI of an N-terminal region mediates its interaction with RNA. The envelope glycoproteins E1 and E2 are highly modified with N-linked sugars, contain intramolecular disulphide bonds and undergo a complex folding pathway that involves several ER-resident chaperones (reviewed in ref. 13). When coexpressed, E1 and E2 are

retained in the ER and form non-covalent heterodimers through determinants in their transmembrane domains.

E2 binds with high affinity to the large extracellular loop of CD81, a tetraspanin that is expressed on a variety of cell types, including hepatocytes⁸⁵. Although this pattern of expression does not explain the hepatotropism of HCV, CD81 is very likely to be involved in mediating HCV entry. Several other candidate HCV receptors have also been identified, including low-density lipoprotein receptor, scavenger receptor class-B type-I (SR-BI), L-SIGN and DC-SIGN¹³. Several surrogate systems have been developed to examine the relevance of these interactions and to study E1/E2 structure and function. These include glycoprotein-dependent cell fusion assays, liposomes reconstituted with E1 and E2, formation of virus-like particles in insect cells, and pseudotyped rhabdoviruses and retroviruses (reviewed in ref. 13). Although each of these systems had merit, the use of retrovirus pseudoparticles (HCVpp) has provided the most insight into HCV entry. This method takes advantage of the fact that retroviruses, which bud from the plasma membrane, frequently and nonspecifically incorporate cell surface proteins into the viral membrane. Thus HCVpp are retroviral particles that are dependent on HCV glycoproteins to deliver a reporter gene encoded within the retrovirus genome.

HCVpp can be neutralized with antibodies against E2 or immune sera, confirming their dependence on E2 and showing their use in following the kinetics and specificity of the humoral immune response^{86,87}. HCVpp infect primary human hepatocytes and a variety of human hepatic cell lines, and their entry is CD81 dependent^{88–91}. CD81 expression is not sufficient for HCV entry into non-hepatic cells, suggesting the existence of one or more unidentified molecules required for HCV entry and hepatotropism. High-density lipoproteins (HDL) and apolipoprotein C1, a component of HDL, enhance HCVpp infectivity in an SR-BI-dependent manner^{87,92}. It is not yet clear whether these observations are related to the association of HCV with serum lipoproteins.

Despite these advances, the growth of authentic HCV in cell culture has remained elusive. Although full-length genomes harbouring adaptive mutations replicated efficiently in Huh-7 cells and expressed the structural proteins, infectious particles were not released^{61,93,94}. This led to the idea that Huh-7 cells might be unable to support HCV particle assembly or release. However, culture-adaptive changes were also found to be lethal or highly attenuating for replication in chimpanzees⁹⁵. This suggested that adaptive mutations promoting efficient RNA replication in cell culture might preclude production of infectious particles. In support of this idea, Pietschmann *et al.* found that full-length genomes lacking adaptive mutations replicated poorly but nevertheless released core protein and HCV RNA into the cell culture medium⁹⁶. In contrast, genomes with highly adaptive mutations replicated efficiently but failed

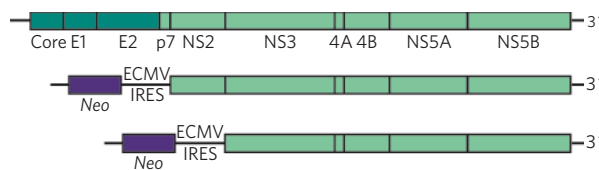


Figure 4 | Cell culture replication systems. The basic design of subgenomic replicons.

to secrete HCV RNA and core protein. Thus, there was a dilemma: in cell culture, adaptive mutations were required for replication but were deleterious for virus production; *in vivo*, cell adaptive mutations were deleterious and virus produced *in vivo* was non-infectious in cell culture. Enter JFH-1: a genotype 2a HCV isolate obtained from a patient in Japan with fulminant hepatitis⁹⁷. For reasons that are not understood, subgenomic replicons derived from JFH-1 cDNA do not require adaptive mutations for efficient replication in cell culture. Wakita *et al.* recently demonstrated that the full-length JFH-1 genome produces infectious particles in cell culture, although the titres were moderate (ref. 98).

Similarly, we constructed a chimaeric full-length genome using the JFH-1 replicase and the core–NS2 region from a related genotype 2a strain, J6. This genome replicated in cell culture and produced robust levels of infectious virus (HCVcc), nearly 10^5 infectious units ml^{-1} within 48 h in Huh-7.5 cells, which are highly permissive for HCV RNA replication⁹⁹. Another group has found that full-length JFH-1 can also reach high titres when propagated in an Huh-7.5 subline¹⁰⁰. Thus, the JFH-1 replicon can support efficient production of infectious HCV in cell culture. It is not yet clear why this particular genome is capable of replicating without adaptive changes, or how adaptive changes preclude infectious particle production. However, one possible explanation stems from the three coordinated yet distinct processes facing HCV viral RNA: translation, replication and packaging. Cell culture adaptive changes, selecting for efficient and persistent RNA replication, may shift the balance towards these first two processes at the expense of liberating genome RNA for virion assembly.

Outlook

It is clearly an exciting time in HCV research, and rapid progress should be made now that complete cell culture systems are available. The processes of HCV entry, replication and virion production can be further dissected with genetic and biochemical approaches, and should reveal information that facilitates the development of specific antivirals that target each stage in the virus life cycle. If the determinants of JFH-1 that permit efficient replication and virus production can be mapped, it may be possible to extend the cell culture systems to include other virus isolates as well. An important question is whether JFH-1-derived viruses grown *in vivo* will retain their infectivity in cell culture. Furthermore, once cellular determinants of HCV tropism are better understood, it might be possible to engineer improved small-animal models of HCV infection and pathogenesis. These approaches will undoubtedly reveal new and surprising aspects of HCV replication, and arm us with better strategies to combat HCV infection and eradicate HCV-associated disease. ■

- Feinstone, S. M., Kapikian, A. Z., Purcell, R. H., Alter, H. J. & Holland, P. V. Transfusion-associated hepatitis not due to viral hepatitis type A or B. *N. Engl. J. Med.* **292**, 767–770 (1975).
- Choo, Q.-L. *et al.* Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**, 359–362 (1989).
- Neumann, A. U. *et al.* Hepatitis C viral dynamics *in vivo* and the antiviral efficacy of interferon- α therapy. *Science* **282**, 103–107 (1998).
- Hoofnagle, J. H. Course and outcome of hepatitis C. *Hepatology* **36**, S21–S29 (2002).
- Lindenbach, B. D. & Rice, C. M. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 991–1041 (Lippincott-Raven, Philadelphia, 2001).
- Deforges, S. *et al.* Expression of hepatitis C virus proteins in epithelial intestinal cells *in vivo*. *J. Gen. Virol.* **85**, 2515–2523 (2004).

- Forton, D. M., Karayiannis, P., Mahmud, N., Taylor-Robinson, S. D. & Thomas, H. C. Identification of unique hepatitis C virus quasispecies in the central nervous system and comparative analysis of internal translational efficiency of brain, liver, and serum variants. *J. Virol.* **78**, 5170–5183 (2004).
- Bigger, C. B., Brasky, K. M. & Lanford, R. E. DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. *J. Virol.* **75**, 7059–7066 (2001).
- Bigger, C. B. *et al.* Intrahepatic gene expression during chronic hepatitis C virus infection in chimpanzees. *J. Virol.* **78**, 13779–13792 (2004).
- Su, A. I. *et al.* Genomic analysis of the host response to hepatitis C virus infection. *Proc. Natl Acad. Sci. USA* **99**, 15669–15674 (2002).
- Thimme, R. *et al.* Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease. *Proc. Natl Acad. Sci. USA* **99**, 15661–15668 (2002).
- Simmonds, P. Genetic diversity and evolution of hepatitis C virus—15 years on. *J. Gen. Virol.* **85**, 3173–3188 (2004).
- Bartenschlager, R., Frese, M. & Pietschmann, T. Novel insights into hepatitis C virus replication and persistence. *Adv. Virus Res.* **63**, 71–180 (2004).
- Penin, F., Dubuisson, J., Rey, F. A., Moradpour, D. & Pawlowsky, J. M. Structural biology of hepatitis C virus. *Hepatology* **39**, 5–19 (2004).
- Spahn, C. M. *et al.* Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science* **291**, 1959–1962 (2001).
- Ji, H., Fraser, C. S., Yu, Y., Leary, J. & Doudna, J. A. Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proc. Natl Acad. Sci. USA* **101**, 16990–16995 (2004).
- Otto, G. A. & Puglisi, J. D. The pathway of HCV IRES-mediated translation initiation. *Cell* **119**, 369–380 (2004).
- Griffin, S. D. *et al.* The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. *FEBS Lett.* **535**, 34–38 (2003).
- Pavlovic, D. *et al.* The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. *Proc. Natl Acad. Sci. USA* **100**, 6104–6108 (2003).
- Branch, A. D., Stump, D. D., Gutierrez, J. A., Eng, F. & Walewski, J. L. The hepatitis C virus alternate reading frame (ARF) and its family of novel products: the alternate reading frame protein/F-protein, the double-frameshift protein, and others. *Semin. Liver Dis.* **25**, 105–117 (2005).
- Failla, C., Tomei, L. & De Francesco, R. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *J. Virol.* **68**, 3753–3760 (1994).
- Bartenschlager, R., Ahlborn-Laake, L., Mous, J. & Jacobsen, H. Kinetic and structural analyses of hepatitis C virus polyprotein processing. *J. Virol.* **68**, 5045–5055 (1994).
- Lin, C., Thomson, J. A. & Rice, C. M. A central region in the hepatitis C virus NS4A protein allows formation of an active NS3–NS4A serine proteinase complex *in vivo* and *in vitro*. *J. Virol.* **69**, 4373–4380 (1995).
- Wölk, B. *et al.* Subcellular localization, stability, and trans-cleavage competence of the hepatitis C virus NS3–NS4A complex expressed in tetracycline-regulated cell lines. *J. Virol.* **74**, 2293–2304 (2000).
- Tai, C.-L., Chi, W.-K., Chen, D.-S. & Hwang, L.-H. The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *J. Virol.* **70**, 8477–8484 (1996).
- Levin, M. K., Gurjar, M. & Patel, S. S. A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase. *Nature Struct. Mol. Biol.* **12**, 429–435 (2005).
- Serebrov, V. & Pyle, A. M. Periodic cycles of RNA unwinding and pausing by hepatitis C virus NS3 helicase. *Nature* **430**, 476–480 (2004).
- Frick, D. N., Rypma, R. S., Lam, A. M. & Gu, B. The nonstructural protein 3 protease/helicase requires an intact protease domain to unwind duplex RNA efficiently. *J. Biol. Chem.* **279**, 1269–1280 (2004).
- Pang, P. S., Jankowsky, E., Planet, P. J. & Pyle, A. M. The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. *EMBO J.* **21**, 1168–1176 (2002).
- He, Y. & Katze, M. G. To interfere and to anti-interfere: the interplay between hepatitis C virus and interferon. *Viral Immunol.* **15**, 95–119 (2002).
- Macdonald, A. & Harris, M. Hepatitis C virus NS5A: tales of a promiscuous protein. *J. Gen. Virol.* **85**, 2485–2502 (2004).
- Tellinghuisen, T. L. & Rice, C. M. Interaction between hepatitis C virus proteins and host cell factors. *Curr. Opin. Microbiol.* **5**, 419–427 (2002).
- Katze, M. G. *et al.* Ser²⁹⁴ is a highly conserved major phosphorylation site of the hepatitis C virus nonstructural protein NS5A. *Virology* **278**, 501–513 (2000).
- Reed, K. E. & Rice, C. M. Identification of the major phosphorylation site of the hepatitis C virus H strain NS5A protein as serine 2321. *J. Biol. Chem.* **274**, 28011–28018 (1999).
- Coito, C., Diamond, D. L., Neddermann, P., Korth, M. J. & Katze, M. G. High-throughput screening of the yeast kinome: identification of human serine/threonine protein kinases that phosphorylate the hepatitis C virus NS5A protein. *J. Virol.* **78**, 3502–3513 (2004).
- Ide, Y., Tanimoto, A., Sasaguri, Y. & Padmanabhan, R. Hepatitis C virus NS5A protein is phosphorylated *in vitro* by a stably bound protein kinase from HeLa cells and by cAMP-dependent protein kinase A- α catalytic subunit. *Gene* **201**, 151–158 (1997).
- Kim, J., Lee, D. & Choe, J. Hepatitis C virus NS5A protein is phosphorylated by casein kinase II. *Biochem. Biophys. Res. Commun.* **257**, 777–781 (1999).
- Reed, K. E., Xu, J. & Rice, C. M. Phosphorylation of the hepatitis C virus NS5A protein *in vitro* and *in vivo*: properties of the NS5A-associated kinase. *J. Virol.* **71**, 7187–7197 (1997).
- Brass, V. *et al.* An amino-terminal amphipathic α -helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *J. Biol. Chem.* **277**, 8130–8139 (2002).
- Tellinghuisen, T. L., Marcotrigiano, J., Goralenya, A. E. & Rice, C. M. The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J. Biol. Chem.* **279**, 48576–48587 (2004).
- Tellinghuisen, T. L., Marcotrigiano, J. & Rice, C. M. Structure of the zinc-binding domain of an essential replicase component of hepatitis C virus. *Nature* **435**, 374–379 (2005).
- Appel, N., Pietschmann, T. & Bartenschlager, R. Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *J. Virol.* **79**, 3187–3194 (2005).

43. Moradpour, D. *et al.* Insertion of green fluorescent protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus replication complexes. *J. Virol.* **78**, 7400–7409 (2004).
44. Hong, Z. *et al.* A novel mechanism to ensure terminal initiation by hepatitis C virus NS5B polymerase. *Virology* **285**, 6–11 (2001).
45. Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* **410**, 235–240 (2001).
46. van Dijk, A. A., Makeyev, E. V. & Bamford, D. H. Initiation of viral RNA-dependent RNA polymerization. *J. Gen. Virol.* **85**, 1077–1093 (2004).
47. Bressanelli, S., Tomei, L., Rey, F. A. & De Francesco, R. Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *J. Virol.* **76**, 3482–3492 (2002).
48. Schmidt-Mende, J. *et al.* Determinants for membrane association of the hepatitis C virus RNA-dependent RNA polymerase. *J. Biol. Chem.* **276**, 44052–44063 (2001).
49. Wang, Q. M. *et al.* Oligomerization and cooperative RNA synthesis activity of hepatitis C virus RNA-dependent RNA polymerase. *J. Virol.* **76**, 3865–3872 (2002).
50. Kolykhalov, A. A., Feinstone, S. M. & Rice, C. M. Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *J. Virol.* **70**, 3363–3371 (1996).
51. Tanaka, T., Kato, N., Cho, M. J. & Shimotohno, K. A novel sequence found at the 3' terminus of hepatitis C virus genome. *Biochem. Biophys. Res. Commun.* **215**, 744–749 (1995).
52. Kolykhalov, A. A. *et al.* Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA. *Science* **277**, 570–574 (1997).
53. Yanagi, M., Purcell, R. H., Emerson, S. U. & Bukh, J. Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee. *Proc. Natl Acad. Sci. USA* **94**, 8738–8743 (1997).
54. Kolykhalov, A. A., Mihalik, K., Feinstone, S. M. & Rice, C. M. Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication *in vivo*. *J. Virol.* **74**, 2046–2051 (2000).
55. Sakai, A. *et al.* The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. *Proc. Natl Acad. Sci. USA* **100**, 11646–11651 (2003).
56. Yanagi, M., St Claire, M., Emerson, S. U., Purcell, R. H. & Bukh, J. *In vivo* analysis of the 3' untranslated region of the hepatitis C virus after *in vitro* mutagenesis of an infectious cDNA clone. *Proc. Natl Acad. Sci. USA* **96**, 2291–2295 (1999).
57. Forns, X. *et al.* Hepatitis C virus lacking the hypervariable region 1 of the second envelope protein is infectious and causes acute resolving or persistent infection in chimpanzees. *Proc. Natl Acad. Sci. USA* **97**, 13318–13323 (2000).
58. Bukh, J. A critical role for the chimpanzee model in the study of hepatitis C. *Hepatology* **39**, 1469–1475 (2004).
59. Lohmann, V. *et al.* Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**, 110–113 (1999).
60. Blight, K. J., Kolykhalov, A. A. & Rice, C. M. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**, 1972–1974 (2000).
61. Blight, K. J., McKeating, J. A., Marcotrigiano, J. & Rice, C. M. Efficient RNA replication of hepatitis C virus genotype 1a in cell culture. *J. Virol.* **77**, 3181–3190 (2003).
62. Lohmann, V., Hoffmann, S., Herian, U., Penin, F. & Bartenschlager, R. Viral and cellular determinants of hepatitis C virus RNA replication in cell culture. *J. Virol.* **77**, 3007–3019 (2003).
63. Evans, M. J., Rice, C. M. & Goff, S. P. Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *Proc. Natl Acad. Sci. USA* **101**, 13038–13043 (2004).
64. Neddermann, P. *et al.* Reduction of hepatitis C virus NS5A hyperphosphorylation by selective inhibition of cellular kinases activates viral RNA replication in cell culture. *J. Virol.* **78**, 13306–13314 (2004).
65. Gao, L., Aizaki, H., He, J. W. & Lai, M. M. Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. *J. Virol.* **78**, 3480–3488 (2004).
66. Egger, D. *et al.* Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *J. Virol.* **76**, 5974–5984 (2002).
67. Gosert, R. *et al.* Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J. Virol.* **77**, 5487–5492 (2003).
68. Einav, S., Elazar, M., Danieli, T. & Glenn, J. S. A nucleotide binding motif in hepatitis C virus (HCV) NS4B mediates HCV RNA replication. *J. Virol.* **78**, 11288–11295 (2004).
69. Kapadia, S. B. & Chisari, F. V. Hepatitis C virus RNA replication is regulated by host geranylgeranylation and fatty acids. *Proc. Natl Acad. Sci. USA* **102**, 2561–2566 (2005).
70. Ye, J. *et al.* Disruption of hepatitis C virus RNA replication through inhibition of host protein geranylgeranylation. *Proc. Natl Acad. Sci. USA* **100**, 15865–15870 (2003).
71. Wang, C. *et al.* Identification of FBL2 as a geranylated cellular protein required for hepatitis C virus RNA replication. *Mol. Cell* **18**, 425–434 (2005).
72. Friebe, P., Lohmann, V., Krieger, N. & Bartenschlager, R. Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J. Virol.* **75**, 12047–12057 (2001).
73. Kim, Y. K., Kim, C. S., Lee, S. H. & Jang, S. K. Domains I and II in the 5' nontranslated region of the HCV genome are required for RNA replication. *Biochem. Biophys. Res. Commun.* **290**, 105–112 (2002).
74. Luo, G., Xin, S. & Cai, Z. Role of the 5'-proximal stem-loop structure of the 5' untranslated region in replication and translation of hepatitis C virus RNA. *J. Virol.* **77**, 3312–3318 (2003).
75. Reusken, C. B., Dalebout, T. J., Erlich, P., Bredenbeek, P. J. & Spaan, W. J. Analysis of hepatitis C virus/classical swine fever virus chimeric 5' NTRs: sequences within the hepatitis C virus IRES are required for viral RNA replication. *J. Gen. Virol.* **84**, 1761–1769 (2003).
76. Friebe, P. & Bartenschlager, R. Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. *J. Virol.* **76**, 5326–5338 (2002).
77. Yi, M. & Lemon, S. M. 3' nontranslated RNA signals required for replication of hepatitis C virus RNA. *J. Virol.* **77**, 3557–3568 (2003).
78. You, S., Stump, D. D., Branch, A. D. & Rice, C. M. A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication. *J. Virol.* **78**, 1352–1366 (2004).
79. Friebe, P., Boudet, J., Simorre, J. P. & Bartenschlager, R. Kissing-loop interaction in the 3' end of the hepatitis C virus genome essential for RNA replication. *J. Virol.* **79**, 380–392 (2005).
80. Miyazaki, Y. *et al.* Hepatitis C virus non-structural proteins in the probable membranous compartment function in viral genome replication. *J. Biol. Chem.* **278**, 50301–50308 (2003).
81. Yang, G. *et al.* Newly synthesized hepatitis C virus replicon RNA is protected from nuclease activity by a protease-sensitive factor(s). *J. Virol.* **78**, 10202–10205 (2004).
82. Barba, G. *et al.* Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets. *Proc. Natl Acad. Sci. USA* **94**, 1200–1205 (1997).
83. Moradpour, D., Englert, C., Wakita, T. & Wands, J. R. Characterization of cell lines allowing tightly regulated expression of hepatitis C virus core protein. *Virology* **222**, 51–63 (1996).
84. Yasui, K. *et al.* The native form and maturation process of hepatitis C virus core protein. *J. Virol.* **72**, 6048–6055 (1998).
85. Pileri, P. *et al.* Binding of hepatitis C virus to CD81. *Science* **282**, 938–941 (1998).
86. Logvinoff, C. *et al.* Neutralizing antibody response during acute and chronic hepatitis C virus infection. *Proc. Natl Acad. Sci. USA* **101**, 10149–10154 (2004).
87. Meunier, J. C. *et al.* Evidence for cross-genotype neutralization of hepatitis C virus pseudo-particles and enhancement of infectivity by apolipoprotein C1. *Proc. Natl Acad. Sci. USA* **102**, 4560–4565 (2005).
88. Cormier, E. G. *et al.* CD81 is an entry coreceptor for hepatitis C virus. *Proc. Natl Acad. Sci. USA* **101**, 7270–7274 (2004).
89. Lavillette, D. *et al.* Characterization of host-range and cell entry properties of the major genotypes and subtypes of hepatitis C virus. *Hepatology* **41**, 265–274 (2005).
90. McKeating, J. A. *et al.* Diverse hepatitis C virus glycoproteins mediate viral infection in a CD81-dependent manner. *J. Virol.* **78**, 8496–8505 (2004).
91. Zhang, J. *et al.* CD81 is required for hepatitis C virus glycoprotein-mediated viral infection. *J. Virol.* **78**, 1448–1455 (2004).
92. Voisset, C. *et al.* High density lipoproteins facilitate hepatitis C virus entry through the scavenger receptor class B type I. *J. Biol. Chem.* **280**, 7793–7799 (2005).
93. Blight, K. J., McKeating, J. A. & Rice, C. M. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J. Virol.* **76**, 13001–13014 (2002).
94. Pietschmann, T. *et al.* Persistent and transient replication of full-length hepatitis C virus genomes in cell culture. *J. Virol.* **76**, 4008–4021 (2002).
95. Bukh, J. *et al.* Mutations that permit efficient replication of hepatitis C virus RNA in Huh-7 cells prevent productive replication in chimpanzees. *Proc. Natl Acad. Sci. USA* **99**, 14416–14421 (2002).
96. Pietschmann, T. & Bartenschlager, R. In Proc. 11th International Symposium on Hepatitis C Virus and Related Viruses (Heidelberg, Germany, 2004).
97. Kato, T. *et al.* Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology* **125**, 1808–1817 (2003).
98. Wakita, T. *et al.* Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Med.* **11**, 791–796 (2005).
99. Lindenbach, B. D. *et al.* Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623–626 (2005).
100. Zhong, J. *et al.* Robust hepatitis C virus infection *in vitro*. *Proc. Natl Acad. Sci. USA* **102**, 9294–9299 (2005).

Acknowledgements We thank our colleagues for many helpful discussions, and in particular T. Tellinghuisen, J. Marcotrigiano, I. Lorenz, T. Pietschmann and R. Bartenschlager for providing data before publication; and J. Bloom, T. Tellinghuisen, M. Evans and I. Lorenz for comments on the manuscript. Work in our laboratory is supported by the US Public Health Service under grants from the NIH to C.M.R., and the Greenberg Medical Research Institute. B.D.L. is a recipient of the NIH Howard Temin Award.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to C.M.R. (ricec@rockefeller.edu).

Evasion of intracellular host defence by hepatitis C virus

Michael Gale Jr¹ and Eileen M. Foy¹

Viral infection of mammalian cells rapidly triggers intracellular signalling events leading to interferon α/β production and a cellular antiviral state. This 'host response' is our first line of immune defence against infection as it imposes several barriers to viral replication and spread. Hepatitis C virus (HCV) evades the host response through a complex combination of processes that include signalling interference, effector modulation and continual viral genetic variation. These evasion strategies support persistent infection and the spread of HCV. Defining the molecular mechanisms by which HCV regulates the host response is of crucial importance and may reveal targets for novel therapeutic strategies.

Hepatitis C virus (HCV) is remarkably successful. It typically produces a persistent infection that, unless interrupted by interferon (IFN)-based therapy, will continue for the lifetime of the individual and present vast opportunities for further transmission within the human population. This success is linked to an ability of HCV to evade and antagonize the immune response of the host and to resist the antiviral actions of IFN therapy. Until recently, native HCV could not be adequately propagated in cultured cells to support molecular studies of the virus–host relationship. Insight into its evasion strategies has come from studies of model systems and human patients. These studies have revealed several levels of immune regulation and evasion conferred by HCV protein products.

This review will explore the mechanisms by which HCV triggers, controls and evades antiviral defences directly within the infected hepatocyte and hepatic tissue to support HCV replication and persistence.

The host response to infection

Virus infection initiates a series of intracellular events that culminate in the generation of an antiviral state directly within the infected cell and indirectly within the surrounding tissue. To replicate and spread successfully, viruses direct various strategies to evade host defences¹. In recent years much has been learned about the molecular mechanisms by which viruses trigger and regulate these antiviral processes, referred to here as the 'host response'. It is the hepatic host response that imposes initial immune defences against HCV infection. The host response is triggered when a pathogen-associated molecular pattern (PAMP) presented by the infecting virus is recognized and engaged by specific PAMP receptor factors expressed in the host cell, initiating signals that ultimately induce the expression of antiviral effector genes². For RNA viruses, protein and nucleic acid products of infection or replication, including single-stranded (ss) or double-stranded (ds) RNA and polyuridine signatures, have been identified as viral PAMPs and are each engaged by specific Toll-like receptors (TLRs) or nucleic-acid-binding proteins that serve as PAMP receptors (Fig. 1)^{3,4}. The viral RNA of HCV contains each of these PAMP signatures and is sufficient to trigger the host response when introduced into naive cells^{5,6}. In hepatocytes (the target cell of HCV infection), independent pathways of retinoic-acid-inducible gene I (RIG-I) and TLR3 signalling comprise two major pathways of host defence triggering by dsRNA^{6–8}.

The largest effect of PAMP receptor engagement is the activation of latent cellular transcription factors that mediate the rapid onset of gene expression, thus marking the immediate-early phase of the host response⁹. Interferon regulatory factor (IRF)-3 (refs 10, 11) and nuclear factor κ B (NF- κ B)¹² figure largely in this response. IRF-3 and NF- κ B are activated through viral PAMP-responsive signalling cascades that culminate in their nuclear translocation and transcription effector actions. IRF-5 and IRF-7 have also been implicated as direct transcription effectors of viral PAMP signalling^{13,14}, although the events that confer their activation are incompletely defined and their role in HCV infection is not known. Parallel processes that activate ATF-2 and direct chromatin remodelling result in the assembly of an enhanceosome complex with IRF-3 and NF- κ B on the IFN- β promoter, leading to a transcriptional response that produces secreted IFN- β from the infected cell². NF- κ B is also involved in inducing the expression of chemokines and proinflammatory cytokines that function in parallel with IFN to mediate the inflammatory response to HCV¹⁵. Secreted IFN- β engages the local tissue through autocrine and paracrine processes of binding the IFN- α/β receptors. This results in activation of the Jak–STAT pathway, in which the receptor-associated Jak and Tyk1 protein kinases catalyse the phosphorylation of signal transducer and activator of transcription (STAT) proteins on critical serine and tyrosine residues. This confers STAT activation and stable association with IRF-9. The resulting IFN-stimulated gene factor-3 (ISGF3) transcription factor complex localizes to the cell nucleus, where it binds to the IFN-stimulated response element (ISRE) within the promoter/enhancer region of IFN-stimulated genes (ISGs). Jak–STAT signalling leads to a second and later wave of transcriptional activity marking ISG expression in the infected cell.

ISGs are the genetic effectors of the host response to virus infection, and the human genome encodes hundreds of ISGs¹⁶. ISG products impart cell or viral-regulatory functions that limit HCV replication through processes that include disruption of viral RNA translation and inhibition of antigenomic strand RNA synthesis^{17–20}. The paracrine effects of IFN- β induce ISG expression within the neighbouring uninfected cells of the local tissue, inducing an antiviral state that limits cell-to-cell virus spread. Many PAMP receptors and their constituent signalling partners are ISGs, and although expressed basally at a low level that facilitates surveillance, their levels increase markedly after

¹Department of Microbiology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9048, USA.



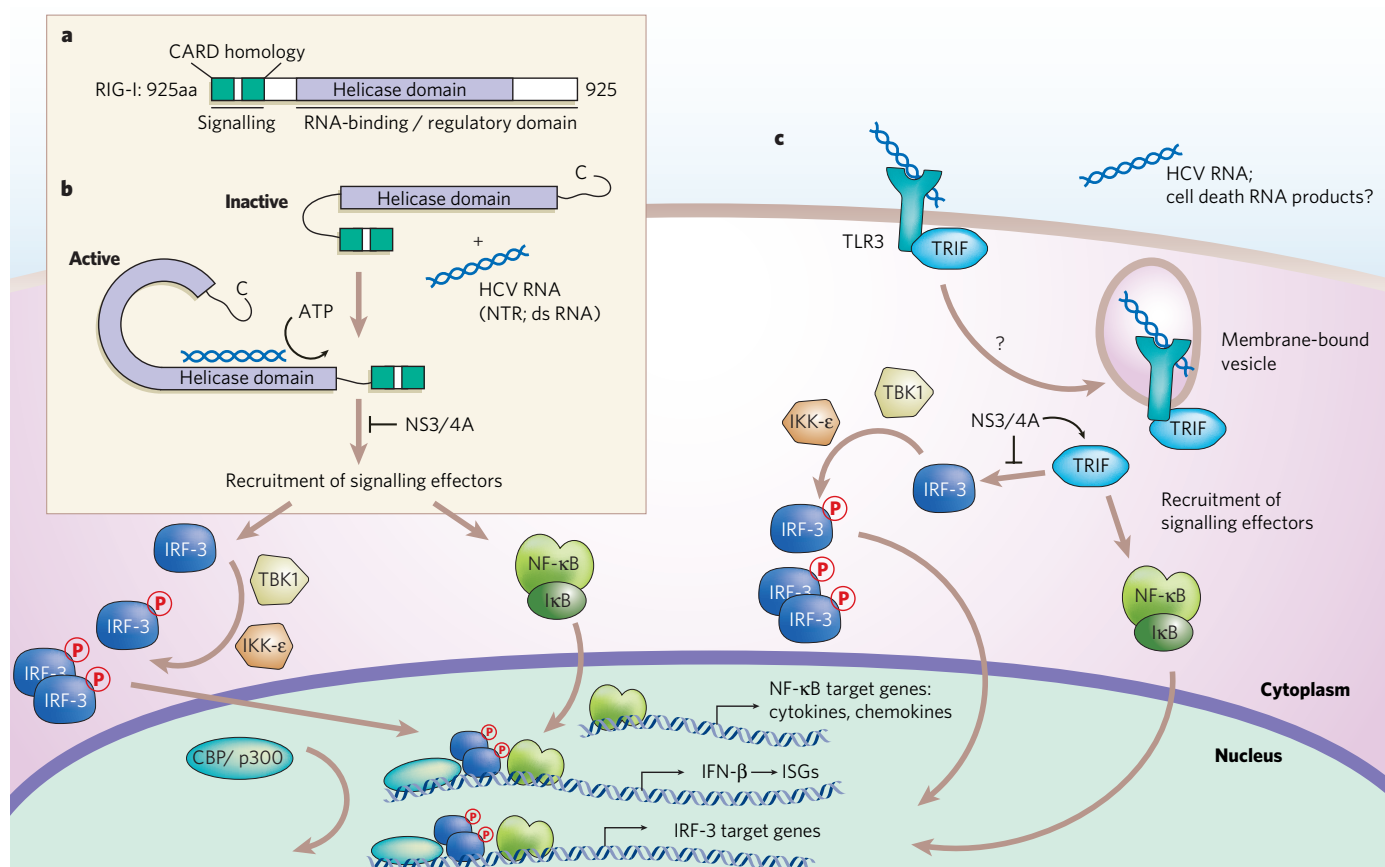


Figure 2 | Triggering IRF-3 activation by HCV through RIG-I or TLR3, and signalling control by NS3/4A. **a**, The domain structure of RIG-I includes N-terminal tandem caspase activation and recruitment domain (CARD) homology regions followed by a C-terminal DEx/D box RNA helicase domain⁷. The RIG-I CARD domains signal IRF-3 and NF-κB activation. The RIG-I helicase domain binds the HCV RNA PAMP⁶. **b**, In the absence of RNA PAMP binding, the RIG-I helicase domain suppresses the signalling actions of the CARD domains, most probably by mediating an autoinhibitory conformation. RNA PAMP binding during HCV RNA replication²¹ is thought to result in an open conformation that permits CARD signalling and probably involves recruitment of signalling factors that direct the phosphorylation and activation of IRF-3. Other factors signal the parallel release of NF-κB from its inhibitor (IκB). CBP, CREB-binding protein. **c**, PAMP signalling through TLR3 is thought to initiate as a result of RNA PAMP binding to the ectodomain of TLR3 that is presented on the cell surface or within membrane-bound cytosolic vesicles. PAMP signalling

through TLR3 requires the TRIF adaptor protein, which signals the downstream phosphorylation and activation of IRF-3 by IKK-ε or TBK1 protein kinases. TRIF also directs the MyD88-independent activation of NF-κB³³. As a result of PAMP signalling through RIG-I or TLR3, the active, nuclear forms of IRF-3 and NF-κB promote the expression of specific target genes that have antiviral or immunomodulatory actions. The NS3/4A protease disrupts viral PAMP signalling^{21,42,46,92}. NS3/4A protease activity cleaves or inactivates one or more signalling components of the RIG-I pathway that are essential for downstream IRF-3 and NF-κB activation (**a**). The shallow and 'featureless' protease substrate-binding cleft of NS3 may accommodate a variety of substrates, allowing NS3/4A to cleave multiple distinct cellular proteins⁴³. Cleavage of TRIF between residue positions 372 and 373 by NS3/4A ablates viral PAMP signalling through the TLR3 pathway. The separate N-terminal and C-terminal proteolytic fragments of TRIF are unstable and signal neither the downstream phosphorylation and activation of IRF-3 nor the activation of NF-κB (**b**)⁴².

cating a complex cross-talk within the host response between parameters of virus, IFN signalling, and the adaptive immune response to HCV infection. ISG expression profiles have also been observed in animals with chronic HCV infection²⁸. These observations demonstrate, first, that the hepatic host response is triggered during HCV infection but is differentially regulated in association with disease course, and second, that in chronic infection HCV can successfully control or evade the host response to persist in the infected cell and hepatic tissue.

Triggering the host response to HCV infection

Although HCV is a ssRNA virus, its genome RNA encodes regions of extensive secondary dsRNA structure that impart potential PAMP signatures, thus presenting the possibility that HCV RNA is recognized and engaged by host-cell PAMP receptors during infection²⁹. Various studies have shown that genome-length or specific subgenomic fragments of HCV RNA are sufficient to trigger IFN-β promoter activation and IFN production when introduced into cultured human hepatoma cells^{5,30}, indicating that during infection these HCV RNA motifs are recognized and engaged by PAMP receptor(s) that trigger the host response⁶.

The nature of at least one HCV RNA PAMP receptor was revealed

through complementation studies of cells with defective host-response signalling programmes⁷, including cells that are highly permissive to HCV RNA replication^{6,31}. This work identified RIG-I as a viral PAMP receptor that binds dsRNA⁷, including dsRNA motifs of the HCV genome⁶, to signal the downstream activation of IRF-3 and NF-κB, thereby inducing IFN-β expression and onset of the host response. RIG-I is a DEx/D-box RNA helicase belonging to a small family of helicases involved in host response signalling^{7,32}. RIG-I contains amino-terminal regions of homology to the caspase activation and recruitment domain (CARD; Fig. 2a). Signalling is mediated by the CARD homology motifs, which direct the downstream activation of IRF-3 and NF-κB through processes independent of TLR3 (Fig. 2b)⁷. TLR3 is a dsRNA PAMP receptor that also signals a host response on engagement of a dsRNA ligand (Fig. 2c)⁷. TLR3 directs the activation of IRF-3 and NF-κB through processes that require the protein Toll-interleukin-1 receptor-domain-containing adaptor inducing IFN-β (TRIF)³³. Virus signalling through RIG-I and TLR3 pathways confers a host response that regulates cellular permissiveness for viral replication.

Protein products of virus infection may also stimulate the host response or activate specific components of this response as they accu-

multate in the cell. Expression of the HCV NS5A protein induces cellular stress and signalling pathways that activate STAT3 (ref. 34). STAT3 promotes gene expression through processes that involve the Jak-STAT pathway³⁵. This results in a gene-expression profile that includes ISGs and proinflammatory cytokines that may influence the overall level of HCV RNA replication³⁶. Moreover, the HCV core protein can activate protein kinase R (PKR), a cellular antiviral protein kinase and an effector component of the host response to virus infection³⁷. PKR is an ISG and a dsRNA-binding protein whose RNA-dependent activation results in localized translational suppression and parallel stimulation of NF- κ B and IRF-1 transcription-effector actions³⁸. It is likely that PKR activation by the core protein is associated with the similar ability of the latter to bind RNA³⁹, thereby providing the PKR activator substrate and a possible mechanism of PKR activation during HCV infection. Cell interactions with virus particles may also trigger signalling events that induce IFN production. HCV pseudo-particle binding to dendritic cells has been shown to mediate particle uptake and dendritic cell activation⁴⁰. Because dendritic cells represent a major source of IFN production during viral infection, modulation of their function may influence systemic and/or local IFN signalling and ISG expression⁴¹.

Control and evasion of the host response

The success of HCV in persisting is linked to its overall ability to disrupt the host response and evade antiviral defences. Key sites of HCV control over the host response are found within the PAMP-responsive signalling pathways that impart IRF-3 activation, within the IFN- α/β receptor signalling pathway that confers ISG expression, and at the level of ISG effector protein products¹. The HCV NS3/4A protease functions as an antagonist of virus-induced IRF-3 activation and IFN- β expression through its ability to block RIG-I signalling and to ablate TLR3 signalling by cleaving the TRIF adaptor protein^{21,42,43} (Fig. 2). NS3 is a bifunctional enzyme. Its N-terminal domain encodes a serine protease, and its carboxy-terminal domain encodes an RNA helicase; the latter may support replication by unwinding the viral RNA⁴⁴. The NS3/4A complex constitutes the essential viral protease, which liberates the non-structural proteins from the HCV polyprotein during virus replication⁴⁵. The helicase activity of NS3 is dispensable for the control of IRF-3 activation, but NS3/4A protease activity is required for this regulation⁴⁶. The proteolytic targeting of host factors by NS3/4A as an evasion strategy from host defence was affirmed through pharmacological studies with a peptidomimetic active-site NS3 protease inhibitor. Treatment of cells that express functional NS3/4A alone or in the context of active HCV RNA replication showed that the protease inhibitor effectively removed the blockade to RIG-I and TLR3 signalling imposed by NS3/4A, thereby restoring virus-induced IRF-3 phosphorylation/activation and the activation of NF- κ B^{21,42,46}. This provides pharmacological confirmation that the protease action of NS3/4A is a functional antagonist of the host response induced by dsRNA and viral PAMP signalling.

The viral disruption of RIG-I or TLR3 signalling has many implications. First, this control attenuates two major pathways of IFN production in hepatocytes⁸. Second, many of the components of these pathways, including RIG-I, TLR3, TRIF and the downstream I κ B kinase (IKK)- ϵ kinase (one of the enzymes that can phosphorylate and activate IRF-3)^{47,48}, are responsive to IFN and although expressed at low basal levels their abundance is induced severalfold on exposure of cells to IFN- α/β . The IFN-responsiveness of these factors confers amplification of PAMP signalling action to further enhance the magnitude and duration of the host response. The signalling blockade imposed by NS3/4A breaks this IFN amplification loop (see Table 1)²¹. Third, the MHC components of antigen processing and presentation are themselves ISG products¹⁶, and host-response regulation may effect alterations in antigen presentation, leading to inefficient activation of cytolytic T cells and an inability of the adaptive immune response to clear HCV-infected hepatocytes^{27,49}. Fourth, in addition to its role in host defence, IRF-3 has been ascribed proapoptotic and tumour suppressor functions^{50,51}. In this case, prolonged blocking of IRF-3 function could disrupt these actions, perhaps

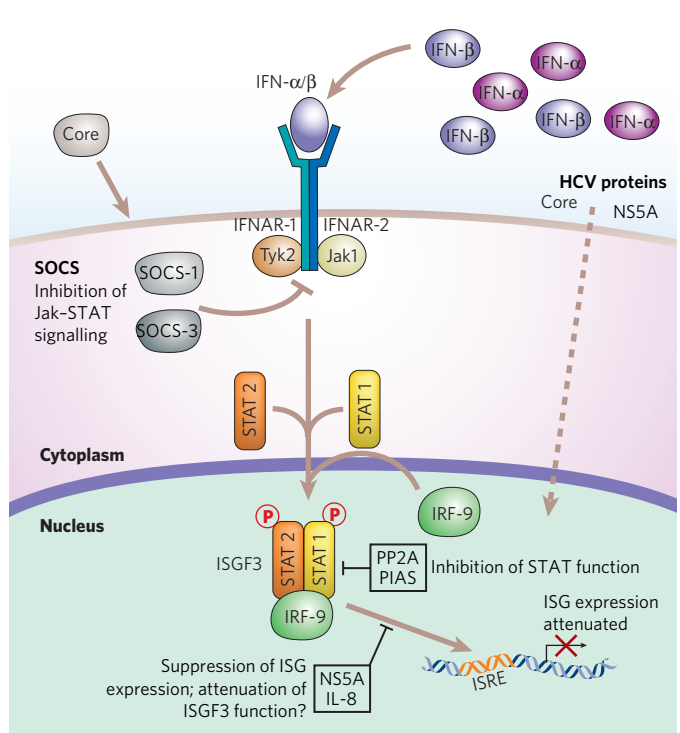


Figure 3 | HCV attenuates IFN signalling through multiple mechanisms.

Receptor signalling by IFN from autocrine/paracrine and therapeutic sources is subject to feedback inhibition by suppressor of cytokine signalling (SOCS) proteins. The HCV core protein has been shown to induce the aberrant expression of SOCS-3, which can suppress Jak-STAT signalling events and block the IFN-induced formation of ISGF3 (ref. 58). HCV protein expression in liver cells is associated with induction of the protein inhibitor of activated STAT (PIAS) expression and concomitant inhibition of STAT function *in vivo*, possibly mediated by protein phosphatase 2A (PP2A) signalling events and STAT demethylation⁵⁶. Patients with chronic HCV infection have exhibited high levels of serum IL-8 (ref. 64). The biological activity of IL-8 interferes with IFN signalling events that catalyse ISGF3 assembly and function⁶². HCV modulation of IFN signalling events attenuates ISG expression, allowing HCV to evade the antiviral actions of the host response and IFN therapy.

to render a tumorigenic potential to infected cells, thus providing a possible biochemical link between chronic HCV and hepatocellular carcinoma⁵². Last, the blockade of virus-induced NF- κ B activity regulates the expression of a variety of chemokines and cytokine genes whose expression is dependent on NF- κ B²¹. Among these is interleukin (IL)-1, which mediates antiviral actions against HCV⁵³. Viral control of NF- κ B may therefore contribute to the broader systemic immune defects and enhanced permissiveness for HCV infection.

Regulation of IFN signalling

Local IFN production in hepatic tissue is likely to influence HCV replication and may impart antiviral effects that contribute to the resolution of acute infection²⁶. The overall low response rate of HCV (particularly genotype 1 HCV) to IFN therapy²⁴ indicates that HCV can evade or resist IFN actions *in vivo*, both locally in the context of a hepatic host response and more globally in the context of IFN therapy (Fig. 3). Assessment of IFN- α/β receptor signalling processes has revealed mechanisms by which HCV proteins can antagonize IFN signalling. HCV protein expression in general has been associated with the inhibition of STAT1 function independently of STAT tyrosine phosphorylation^{54,55}. This has been attributed in part to the expression of high levels of protein phosphatase 2A within HCV-infected liver tissue, which may signal STAT1 hypomethylation and inactivation⁵⁶. Expression of the HCV core protein has been associated with increased expression levels of suppressor of cytokine signalling (SOCS)-3 in cultured cells⁵⁷. The SOCS proteins are best known for their role as negative regulators and inhibitors of Jak-STAT signalling, where they

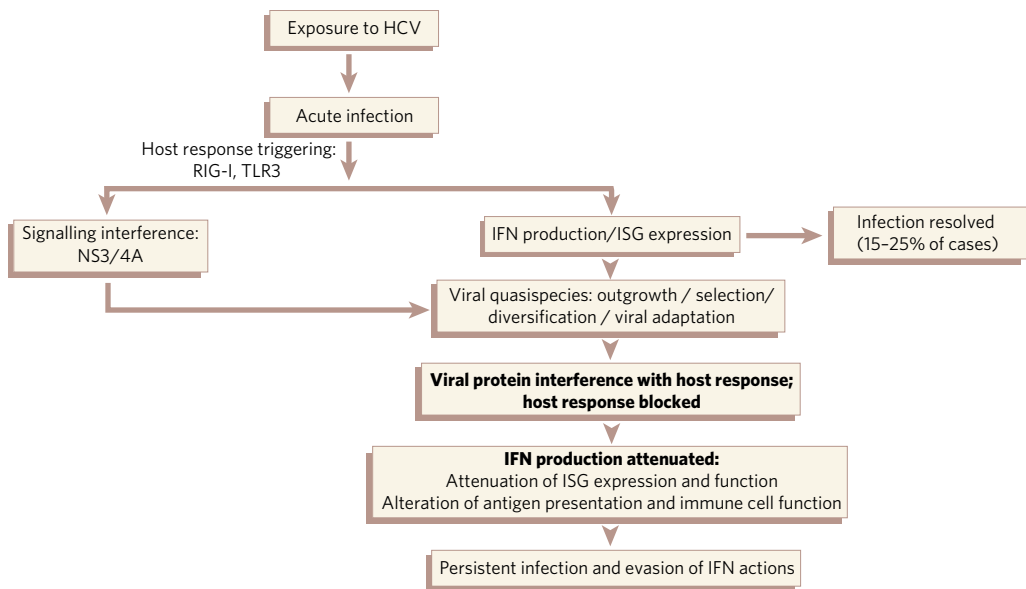


Figure 4 | HCV–host interactions regulate the host response and affect the outcome of HCV infection. A flow diagram (described in the text) is shown in which virus–host interactions within the host response to HCV infection define the outcome from acute exposure to HCV.

mediate a classic negative feedback loop on IFN- α/β receptor signalling events⁵⁸. Induction of SOCS-3 expression by the HCV core protein could impart evasion from IFN actions, but the overall role of SOCS-3 in HCV infection is not known. The actions of IFN are pleiotropic, both at the signalling level and the ISG response, and it is most likely that HCV evades IFN effects through multiple strategies, including possible disruption of non-canonical IFN signalling pathways⁵⁹.

Regulation of ISG expression or function

HCV evasion of the host response includes various strategies directed by viral proteins to control ISG expression or function (Table 1). The HCV NS5A protein has been identified as an IFN antagonist, and expression of NS5A alone can suppress IFN- α actions sufficiently to rescue the replication of an IFN-sensitive virus in cultured cells⁶⁰. Functional genomics analyses have shown that NS5A expression confers a general attenuation of ISG expression⁶¹. A explanation for this comes from the observation that NS5A can induce IL-8 expression and secretion. IL-8 is a proinflammatory chemokine whose actions interfere with IFN⁶². NS5A stimulates IL-8 production through transactivation of the IL-8 promoter⁶³, and serum IL-8 levels have been found elevated in patients with chronic hepatitis C⁶⁴. The mechanisms by which IL8 antagonizes IFN actions are not known but probably involve an end result of altering ISG expression. The HCV NS5A and E2 proteins of HCV are both inhibitors of

PKR^{65–67}. Inhibition of PKR may allow HCV to evade in part the translational-suppressive actions of IFN and the PKR-dependent signalling processes that amplify the host response to infection¹. However, this regulation is not universal and is subject to alteration through viral genetic variation (see below)⁶⁸, indicating that evasion of PKR-independent processes of ISG function contribute to HCV escape from IFN action. The ISG56 product, p56, imparts translational suppressive actions of IFN on HCV RNA replication¹⁷, and reduced levels of ISG56 expression have been associated with IFN resistance of HCV RNA replication *in vitro*⁶⁹. Further examination of HCV interactions with the IFN-induced 2',5'-oligoadenylate synthetase (OAS)/RNase L pathway have revealed that HCV proteins also interact with this pathway⁷⁰, and that once activated the pathway directs the capacity of RNase L to cleave HCV genomic RNA into non-functional nucleolytic products⁷¹. Genetic studies have revealed that RNase L preferentially cleaves HCV RNA only at certain UU and UA dinucleotide sites⁷². Genotype 1 HCV sequences in general have fewer RNase L cleavage sites than HCV genotypes 2 or 3 (ref. 72). This may provide a genetic basis for HCV 1a and 1b resistance to IFN therapy (Table 1).

Viral genetic variation and the host response to infection

As with all RNA viruses, the viral polymerase of HCV lacks a proof-reading function. In the course of persistent infection, error-prone

Table 1 ISG regulation by HCV			
Viral strategy	Mechanism of action	Implications	References
IL-8 induction	NS5A induces IL-8 production through processes involving NF- κ B and AP-1 transcription factor activation	Attenuates ISG expression	63
Induction of SOCS expression	The HCV core protein can induce expression of SOCS1 and SOCS3	Blocks Jak-STAT signalling action through the IFN- α/β receptor	57
PKR inhibition	NS5A and E2 proteins bind PKR and inhibit its catalytic activity	Disruption of PKR-dependent translational control and signalling actions	65,66,68
IRF-1 regulation	NS5A blocks dsRNA-induced IRF-1 action through inhibition of PKR signalling	Relieves IRF-1 suppression of HCV RNA replication	93,94
Evasion of 2',5' OAS/RNase L pathway	HCV genome sequence	The HCV genome encodes a paucity of RNase L recognition sites, which allow protection from nucleolytic processing	71,72
Disruption of STAT1 function	HCV proteins	HCV proteins induce PP2A expression and STAT1 hypomethylation to attenuate ISG expression	54,56
Suppression of ISG56 expression	HCV non-structural proteins	<i>In vitro</i> : NS3/4A and non-structural proteins disrupt virus signalling to the ISG56 promoter. Removes the ISG56 block to viral RNA translation	17,69
Regulation of RIG-I signalling	NS3/4A protease blockade of signalling	Blockade of RIG-I signalling breaks an IFN amplification loop that otherwise enhances ISG expression	21
Regulation of TLR3 signalling	NS3/4A protease cleavage of TRIF	Disruption of a TLR3-pathway IFN amplification loop	42

virus replication generates a repertoire of highly related but genetically distinct viral variants or 'quasispecies'. This is most problematic for the infected patient because quasispecies variation affords remarkable adaptive potential to HCV and has been implicated in evasion and control of the host response to infection and differential sensitivity to IFN therapy⁷³. The hostile antiviral host environment may drive the outgrowth of HCV 'evasion variants' from a pre-existing quasispecies pool or through viral genetic adaptation. Indeed, sequencing studies have shown that the resolution of acute HCV infection is associated with an overall reduction in viral quasispecies complexity within the E1 and E2 coding regions of HCV, whereas progression to chronic infection and resistance to IFN therapy is associated with increased viral genetic complexity^{74,75}. This indicates that host immune pressure may drive the outgrowth or selection of viral evasion variants able to persist and resist IFN action. Sequence analysis of the HCV NS5A coding region has similarly identified specific domains that exhibit sequence variation in association with the outcome of IFN therapy. This association has been variable in different patient populations, but recent meta-analyses and long-term follow-up of these studies provide overall support for NS5A sequence variation within a 40-residue 'interferon sensitivity determining region' (ISDR) that is associated with IFN therapy outcome⁷⁶⁻⁷⁸. This region of NS5A encompasses a genetically flexible domain that is a key site of adaptations that influence HCV RNA replication fitness^{79,80}. Thus, ISDR variation may affect the host response to infection indirectly by altering replication efficiency and the abundance of viral proteins available for interaction with and regulation of host response effectors.

Exogenous induction of antiviral hepatic defences

Multiple studies provide molecular evidence for a clear absence or only a low level of IFN- α/β gene expression in hepatocytes of patients with chronic HCV⁸¹. The expression of IFN- β and various IFN- α subtypes is dependent on virus activation of IRF-3 (ref. 2), and the lack of IFN- α/β gene expression within the HCV-infected liver provides indirect evidence that HCV imposes a blockade to IRF-3 activation *in vivo*. This may explain why some patients with chronic infection do not express significant levels of hepatic ISGs, but it fails to explain why others exhibit broad and abundant ISG expression despite a paucity of IFN- α/β expression in the infected liver. It is notable that hepatic ISG expression has been associated with liver pathology²². This raises the possibility that ISG expression can be induced indirectly as a result of cellular stress from fibrosis and/or cirrhosis, or is induced through TLR engagement exogenously by extracellular products of damaged tissue or viral replication. The former possibility is indicated by cell-culture studies in which stress-induced cytokines, including TNF- α and IL-1, triggered signalling crosstalk to activate IRF-1 and derive a level of IFN- β production⁸². In the context of chronic HCV, the latter possibility could occur through TLR3 engagement of viral or host RNA products by hepatocytes and surrounding cells that are not infected and remain competent to signal an ISG response. Exogenous/extrahepatic immune effector cells that infiltrate the liver, including IFN-producing macrophages and dendritic cells⁸³, may also contribute to hepatic ISG expression. By this model, hepatic ISG levels would vary with the composition and extent of immune cell infiltration, which has been observed^{22,27}. Secretion of IFN- γ by hepatic effector T cells and NK cells also contributes a level of ISG expression partly redundant with the ISGs induced during the IFN- α/β response¹⁶. IFN- γ exerts antiviral effects on HCV RNA replication⁸⁴. This response probably has a role in controlling HCV infection²⁷.

A current model

Studies defining the viral induction, evasion and control of the host response to HCV collectively provide a model of virus-host interactions and viral adaptation that form a foundation for chronic infection (Fig. 4). Transmission of HCV from a source individual and infection of a recipient host present enormous pressure for the virus to adapt to the new host environment and to control the host response to infec-

tion. The transmission event results in an acute infection that involves viral regulation of the host response through RIG-I, TLR3 and other virus-responsive signalling pathways within the infected hepatocyte^{6,8}. Highly fit variants of HCV will mediate signalling interference, in which the NS3/4A protease will block RIG-I and TLR3 signalling pathways to evade the host response to infection and viral RNA replication. Genetic distinctions between virus strains and viral genotypes are likely to impart differential levels of control and activation of this response^{46,68,72}, and during acute infection their activation of the host response will lead to the production of IFN and ISG to mediate an antiviral state in the local hepatic tissue²⁶. About 15–25% of exposures to HCV typically render an acute resolved infection⁸⁵. Thus, if successful, the hepatic host response will provide protection against the replication and spread of HCV. The host response and the ensuing adaptive immune response present pressures that will select for the outgrowth of viral quasi-species that can evade and successfully control the host response and immune defences^{69,75}. HCV-host interactions within RIG-I, TLR3, IFN signalling pathways and ISG pathways, and at other key sites of host defence, serve to control the host response and may attenuate the therapeutic actions of IFN, thus providing a foundation for persistent HCV replication and spread. This model invokes an important role for viral adaptation or quasi-species selection in the successful evasion and control of the host response, and projects a 'foot race' between the virus and the host for control of this response that in most cases the virus will win. The recent development of cell-culture models of HCV infection^{86,87,88} now provides a foundation from which to define the molecular mechanisms and novel sites for therapeutic modulation of the host response controls that regulate the HCV infection and replication cycle. ■

- Katze, M. G., He, Y. & Gale, M. Jr. Viruses and interferon: a fight for supremacy. *Nature Rev. Immunol.* **2**, 675–687 (2002).
- Sen, G. C. Viruses and interferons. *Annu. Rev. Microbiol.* **55**, 255–281 (2001).
- Iwasaki, A. & Medzhitov, R. Toll-like receptor control of the adaptive immune responses. *Nature Immunol.* **5**, 987–995 (2004).
- Cook, D. N., Pisetky, D. S. & Schwartz, D. A. Toll-like receptors in the pathogenesis of human disease. *Nature Immunol.* **5**, 975–979 (2004).
- McCormick, C. J., Challinor, L., Macdonald, A., Rowlands, D. J. & Harris, M. Introduction of replication-competent hepatitis C virus transcripts using a tetracycline-regulable baculovirus delivery system. *J. Gen. Virol.* **85**, 429–439 (2004).
- Sumpter, R. et al. Regulating intracellular anti-viral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *J. Virol.* **79**, 2689–2699 (2005).
- Yoneyama, M. et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
- Li, K., Chen, Z., Kato, N., Gale, M. Jr & Lemon, S. M. Distinct poly-I: C and virus-activated signalling pathways leading to interferon- β production in hepatocytes. *J. Biol. Chem.* **280**, 16739–16747 (2005).
- Malmgaard, L. Induction and regulation of IFNs during viral infections. *J. Interferon Cytokine Res.* **24**, 439–454 (2004).
- Au, W. C., Moore, P. A., Lowther, W., Juang, Y. T. & Pitha, P. M. Identification of a member of the interferon regulatory factor family that binds to the interferon-stimulated response element and activates expression of interferon-induced genes. *Proc. Natl Acad. Sci. USA* **92**, 11657–11661 (1995).
- Lin, R., Heylbroeck, C., Genin, P., Pitha, P. M. & Hiscott, J. Essential role of interferon regulatory factor 3 in direct activation of RANTES chemokine transcription. *Mol. Cell. Biol.* **19**, 959–966 (1999).
- Richmond, A. NF- κ B, chemokine gene transcription and tumour growth. *Nature Rev. Immunol.* **2**, 664–674 (2002).
- Barnes, B. J., Moore, P. A. & Pitha, P. M. Virus-specific activation of a novel interferon regulatory factor, IRF-5, results in the induction of distinct interferon alpha genes. *J. Biol. Chem.* **276**, 23382–23390 (2001).
- Kawai, T. et al. Interferon- α induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nature Immunol.* **5**, 1061–1068 (2004).
- Tai, D. I. et al. Activation of nuclear factor κ B in hepatitis C virus infection: implications for pathogenesis and hepatocarcinogenesis. *Hepatology* **31**, 656–664 (2000).
- Der, S. D., Zhou, A., Williams, B. R. G. & Silverman, R. H. Identification of genes differentially regulated by interferon α , β , or γ using oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* **95**, 15623–15628 (1998).
- Wang, C. et al. Alpha interferon induces distinct translational control programs to suppress hepatitis C virus RNA replication. *J. Virol.* **77**, 3898–3912 (2002).
- Shimazaki, T., Honda, M., Kaneko, S. & Kobayashi, K. Inhibition of internal ribosomal entry site-directed translation of HCV by recombinant IFN- α correlates with a reduced La protein. *Hepatology* **35**, 199–208 (2002).
- Guo, J., Bichko, V. & Seeger, C. Effect of alpha interferon on the hepatitis C virus replication. *J. Virol.* **75**, 8516–8523 (2001).
- Prabhu, R. et al. Interferon α -2b inhibits negative-strand RNA and protein expression from full-length HCV1a infectious clone. *Exp. Mol. Pathol.* **76**, 242–252 (2004).
- Foy, E. et al. Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-

- inducible gene-1 signaling. *Proc. Natl Acad. Sci. USA* **102**, 2986–2991 (2005).
22. Smith, M. W. *et al.* Hepatitis C virus and liver disease: global transcriptional profiling and identification of potential markers. *Hepatology* **38**, 1458–1467 (2003).
 23. Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **435**, 772–777 (2005).
 24. McHutchison, J. G. & Patel, K. Future therapy of hepatitis C. *Hepatology* **36**, S245–S252 (2002).
 25. Walker, C. M. Comparative features of hepatitis C virus infection in humans and chimpanzees. *Springer Semin. Immunopathol.* **19**, 85–98 (1997).
 26. Bigger, C. B., Brasky, K. M. & Lanford, R. E. DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. *J. Virol.* **75**, 7059–7066 (2001).
 27. Su, A. I. *et al.* Genomic analysis of the host response to hepatitis C virus infection. *Proc. Natl Acad. Sci. USA* **99**, 15669–15674 (2002).
 28. Bigger, C. *et al.* Intrahepatic gene expression during chronic hepatitis C virus infection in chimpanzees. *J. Virol.* **78**, 13779–13792 (2004).
 29. Tuplin, A., Wood, J., Evans, D. J., Patel, A. H. & Simmonds, P. Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* **8**, 824–841 (2002).
 30. Fredericksen, B. *et al.* Activation of the interferon- β promoter during hepatitis C virus RNA replication. *Viral Immunol.* **15**, 29–40 (2001).
 31. Blight, K. J., McKeating, J. A. & Rice, C. M. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J. Virol.* **76**, 13001–13014 (2002).
 32. Andrejeva, J. *et al.* The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN- β promoter. *Proc. Natl Acad. Sci. USA* **101**, 17264–17269 (2004).
 33. Yamamoto, M. *et al.* Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science* **301**, 640–643 (2003).
 34. Gong, G., Waris, G., Tanveer, R. & Siddiqui, A. Human hepatitis C virus NS5A protein alters intracellular calcium levels, induces oxidative stress, and activates STAT-3 and NF- κ B. *Proc. Natl Acad. Sci. USA* **98**, 9599–9604 (2001).
 35. Sarcar, B., Ghosh, A. K., Steele, R., Ray, R. & Ray, R. B. Hepatitis C virus NS5A mediated STAT3 activation requires co-operation of Jak1 kinase. *Virology* **322**, 51–60 (2004).
 36. Zhu, H. *et al.* Gene expression associated with interferon α antiviral activity in an HCV replicon cell line. *Hepatology* **37**, 1180–1188 (2003).
 37. Delhem, N. *et al.* Activation of the interferon-inducible protein kinase PKR by hepatocellular carcinoma derived-hepatitis C virus core protein. *Oncogene* **20**, 5836–5845 (2001).
 38. Williams, B. R. PKR, a sentinel kinase for cellular stress. *Oncogene* **18**, 6112–6120 (1999).
 39. Tanaka, Y. *et al.* Selective binding of hepatitis C virus core protein to synthetic oligonucleotides corresponding to the 5' untranslated region of the viral genome. *Virology* **270**, 229–236 (2000).
 40. Barth, H. *et al.* Uptake and presentation of hepatitis C virus-like particles by human dendritic cells. *Blood* **105**, 3605–3614 (2005).
 41. Colonna, M., Trinchieri, G. & Liu, Y. J. Plasmacytoid dendritic cells in immunity. *Nature Immunol.* **5**, 1219–1226 (2004).
 42. Li, K. *et al.* Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc. Natl Acad. Sci. USA* **102**, 2992–2997 (2005).
 43. Ferreón, J. C., Ferreón, A. C., Li, K. & Lemon, S. M. Molecular determinants of TRIF proteolysis mediated by the hepatitis C virus NS3/4A protease. *J. Biol. Chem.* **280**, 20483–20492 (2005).
 44. Reed, K. E. & Rice, C. M. in *Hepatitis C Virus* (ed. Reesink, H. W.) 1–37 (Karger, Basel, 1998).
 45. De Francesco, R. & Steinkühler, C. Structure and function of the hepatitis C virus NS3–NS4A serine proteinase. *Curr. Top. Microbiol. Immunol.* **242**, 149–169 (2000).
 46. Foy, E. *et al.* Regulation of interferon regulatory factor-3 by the hepatitis C virus serine proteinase. *Science* **300**, 1145–1148 (2003).
 47. Sharma, S. *et al.* Triggering the interferon antiviral response through a novel IKK-related pathway. *Science* **300**, 1148–1151 (2003).
 48. Fitzgerald, K. A. *et al.* IKK ϵ and TBK1 are essential components of the IRF3 signaling pathway. *Nature Immunol.* **4**, 491–496 (2003).
 49. Shoukry, N. H., Cawthon, A. G. & Walker, C. M. Cell-mediated immunity and the outcome of hepatitis C virus infection. *Annu. Rev. Microbiol.* **58**, 391–424 (2004).
 50. Duguay, D. *et al.* In vivo interferon regulatory factor 3 tumor suppressor activity in B16 melanoma tumors. *Cancer Res.* **62**, 5148–5152 (2002).
 51. Heylbroeck, C. *et al.* The IRF-3 transcription factor mediates Sendai virus-induced apoptosis. *J. Virol.* **74**, 3781–3792 (2000).
 52. Liang, T. J. & Heller, T. Pathogenesis of hepatitis C-associated hepatocellular carcinoma. *Gastroenterology* **127**, S62–S71 (2004).
 53. Zhu, H. & Liu, C. Interleukin-1 inhibits hepatitis C virus subgenomic RNA replication by activation of extracellular regulated kinase pathway. *J. Virol.* **77**, 5493–5498 (2003).
 54. Heim, M. H., Moradpour, D. & Blum, H. E. Expression of hepatitis C virus proteins inhibits signal transduction through the Jak-STAT pathway. *J. Virol.* **73**, 8469–8475 (1999).
 55. Blindenbacher, A. *et al.* Expression of hepatitis C virus proteins inhibits interferon α signaling in the liver of transgenic mice. *Gastroenterology* **124**, 1465–1475 (2003).
 56. Duong, F. H., Filipowicz, M., Tripodi, M., La Monica, N. & Heim, M. H. Hepatitis C virus inhibits interferon signaling through up-regulation of protein phosphatase 2A. *Gastroenterology* **126**, 263–277 (2004).
 57. Bode, J. G. *et al.* IFN- α antagonistic activity of HCV core protein involves induction of suppressor of cytokine signaling-3. *FASEB J.* **17**, 488–490 (2003).
 58. Alexander, W. S. Suppressors of cytokine signalling (SOCS) in the immune system. *Nature Rev. Immunol.* **2**, 410–416 (2002).
 59. Platanias, L. C. & Fish, E. N. Signaling pathways activated by interferons. *Exp. Hematol.* **27**, 1583–1592 (1999).
 60. Macdonald, A. & Harris, M. Hepatitis C virus NS5A: tales of a promiscuous protein. *J. Gen. Virol.* **85**, 2485–2502 (2004).
 61. Geiss, G. K. *et al.* Gene expression profiling of the cellular transcriptional network regulated by α / β interferon and its partial attenuation by the hepatitis C virus nonstructural 5A protein. *J. Virol.* **77**, 6367–6375 (2003).
 62. Khabar, K. S. *et al.* The alpha chemokine, interleukin 8, inhibits the antiviral action of interferon α . *J. Exp. Med.* **186**, 1077–1085 (1997).
 63. Polyak, S. J. *et al.* Hepatitis C virus nonstructural 5A protein induces interleukin-8, leading to partial inhibition of the interferon-induced antiviral response. *J. Virol.* **75**, 6095–6106 (2001).
 64. Polyak, S. J., Khabar, K. S., Rezeqi, M. & Gretch, D. R. Elevated levels of interleukin-8 in serum are associated with hepatitis C virus infection and resistance to interferon therapy. *J. Virol.* **75**, 6209–6211 (2001).
 65. Taylor, D. R., Shi, S. T., Romano, P. R., Barber, G. N. & Lai, M. M. C. Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science* **285**, 107–110 (1999).
 66. Noguchi, T. *et al.* Effects of mutation in hepatitis C virus nonstructural protein 5A on interferon resistance mediated by inhibition of PKR kinase activity in mammalian cells. *Microbiol. Immunol.* **45**, 829–840 (2001).
 67. Gale, M. Jr *et al.* Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. *Mol. Cell. Biol.* **18**, 5208–5218 (1998).
 68. Gimenez-Barcons, M. *et al.* The oncogenic potential of hepatitis C virus NS5A sequence variants is associated with PKR regulation. *J. Interferon Cytokine Res.* **25**, 152–164 (2005).
 69. Sumpter, R., Wang, C., Foy, E., Loo, Y.-M. & Gale, M. J. Viral evolution and interferon resistance of hepatitis C virus RNA replication in a cell culture model. *J. Virol.* **78**, 11591–11604 (2004).
 70. Taguchi, T. *et al.* Hepatitis C virus NS5A protein interacts with 2',5'-oligoadenylate synthetase and inhibits antiviral activity of IFN in an IFN sensitivity-determining region-independent manner. *J. Gen. Virol.* **85**, 959–969 (2004).
 71. Han, J. Q. & Barton, D. J. Activation and evasion of the antiviral 2',5' oligoadenylate synthetase/ribonuclease L pathway by hepatitis C virus mRNA. *RNA* **8**, 512–525 (2002).
 72. Han, J. Q., Wroblewski, G., Xu, Z., Silverman, R. H. & Barton, D. J. Sensitivity of hepatitis C virus RNA to the antiviral enzyme ribonuclease L is determined by a subset of efficient cleavage sites. *J. Interferon Cytokine Res.* **24**, 664–676 (2004).
 73. Farci, P. Hepatitis C virus. The importance of viral heterogeneity. *Clin. Liver Dis.* **5**, 895–916 (2001).
 74. Farci, P. *et al.* Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. *Proc. Natl Acad. Sci. USA* **99**, 3081–3086 (2002).
 75. Farci, P. *et al.* The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339–344 (2000).
 76. Enomoto, N. *et al.* Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N. Engl. J. Med.* **334**, 77–81 (1996).
 77. Pascu, M. *et al.* Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. *Gut* **53**, 1345–1351 (2004).
 78. Schinkel, J., Spoon, W. J. & Kroes, A. C. Meta-analysis of mutations in the NS5A gene and hepatitis C virus resistance to interferon therapy: uniting discordant conclusions. *Antivir. Ther.* **9**, 275–286 (2004).
 79. Blight, K. J., Kolykhalov, A. A. & Rice, C. M. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**, 1972–1974 (2000).
 80. Appel, N., Pietschmann, T. & Bartenschlager, R. Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *J. Virol.* **79**, 3187–3194 (2005).
 81. Mihm, S. *et al.* Interferon type I gene expression in chronic hepatitis C. *Lab. Invest.* **84**, 1148–1159 (2004).
 82. Fujita, T., Reis, L. F., Watanabe, N., Kimura, Y. & Taniguchi, T. Induction of the transcription factor IRF-1 and interferon- β mRNAs by cytokines and activators of second-messenger pathways. *Proc. Natl Acad. Sci. USA* **86**, 9936–9940 (1986).
 83. Schulz, O. *et al.* Toll-like receptor 3 promotes cross-priming to virus-infected cells. *Nature* **433**, 887–892 (2005).
 84. Frese, M. *et al.* Interferon- γ inhibits replication of subgenomic and genomic hepatitis C virus RNAs. *Hepatology* **35**, 694–703 (2002).
 85. McHutchison, J. G. Understanding hepatitis C. *Am. J. Manag. Care* **10**, S21–S29 (2004).
 86. Zhong, J. *et al.* Robust hepatitis C virus infection *in vitro*. *Proc. Natl Acad. Sci. USA* **102**, 9294–9299 (2005).
 87. Lidenbach, B. D. & Rice, C. M. Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623–626 (2005).
 88. Wakita, T. *et al.* Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Med.* **11**, 791–796 (2005).
 89. Pflugheuer, J. *et al.* Regulation of PKR and IRF-1 during hepatitis C virus RNA replication. *Proc. Natl Acad. Sci. USA* **99**, 4650–4655 (2002).
 90. Kanazawa, N. *et al.* Regulation of hepatitis C virus replication by interferon regulatory factor 1. *J. Virol.* **78**, 9713–9720 (2004).
 91. McWhirter, S. M. *et al.* IFN-regulatory factor 3-dependent gene expression is defective in Tbk1-deficient mouse embryonic fibroblasts. *Proc. Natl Acad. Sci. USA* **101**, 233–238 (2004).
 92. Yoneyama, M. *et al.* Direct triggering of the type I interferon system by virus infection: activation of a transcription factor complex containing IRF-3 and CBP/p300. *EMBO J.* **17**, 1087–1095 (1998).
 93. Au, W. C., Yeow, W. S. & Pitha, P. M. Analysis of functional domains of interferon regulatory factor 7 and its association with IRF-3. *Virology* **280**, 273–282 (2001).
 94. Breiman, A. *et al.* Inhibition of RIG-I-dependent signaling to the interferon pathway during hepatitis C virus expression and restoration of signaling by IKK ϵ . *J. Virol.* **79**, 3969–3978 (2005).

Acknowledgements We thank S. Lemon and members of our laboratory for discussions and critical evaluation of this manuscript. The Gale laboratory is supported by grants from the NIH, the Ellison Medical Foundation and the Burroughs Wellcome Fund. M.G. is the Nancy C. and Jeffrey A. Marcus Scholar in Medical Research in Honor of Dr Bill S. Vowell.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare that they have no competing financial interests. Correspondence and requests for material should be addressed to M.G. (michael.gale@utsouthwestern.edu).

Adaptive immune responses in acute and chronic hepatitis C virus infection

David G. Bowen¹ and Christopher M. Walker^{1,2}

The hepatitis C virus (HCV) persists in the majority of infected individuals and is a significant cause of human illness and death globally. Recent studies have yielded important insights into immunity to HCV, in particular revealing the central role of T cells in viral control and clearance. Other key features of adaptive immune responses remain obscure, including mechanisms by which T cells control HCV replication, the role of antibodies in conferring protection and how cellular and humoral immunity are subverted in persistent infection.

The hepatitis C virus is among the most successful of all persistent human viruses. With a compact RNA genome thought to encode only 11 proteins, HCV persists in up to 70% of those infected by successfully undermining virus-specific immunity while leaving host defences to other infectious agents intact. An estimated 170 million individuals are infected worldwide, and approximately 38,000 new infections occur annually in the United States alone¹. Twenty percent of persistently infected individuals will develop liver cirrhosis, and hepatocellular carcinoma occurs in up to 2.5% (ref. 1). There is as yet no vaccine against HCV. Furthermore, current anti-viral therapy is expensive, relatively toxic and effective in only 50–60 % of patients treated². Understanding adaptive immunity to this virus is crucial for the design of effective strategies to control HCV both in the infected individual and globally.

The outcome of HCV infection is determined within six months of exposure to the virus. Acute infection is often unrecognized because symptoms are usually mild or absent. Initial views of immunity to HCV were therefore largely shaped by studies of chronically infected individuals (reviewed in ref. 3). Recent prospective studies of humans at high risk of HCV exposure and of experimentally infected chimpanzees have provided a more complete picture of adaptive immunity to the virus. In particular, control of acute primary viral replication is associated with expansion of antiviral CD4⁺ (helper) and CD8⁺ (cytotoxic) T cells. Moreover, immunological memory conferred by spontaneous resolution of acute hepatitis C does not protect against reinfection, but does substantially reduce the risk of persistence upon re-exposure. Here we outline features of successful adaptive immune responses to HCV and review current concepts of evasion strategies that might explain defects in humoral and cellular immunity in those individuals who develop persistent infections.

The humoral response to hepatitis C virus infection

Virus-specific antibodies are usually detectable approximately 7–8 weeks after HCV infection⁴. Whether antibodies neutralize HCV infectivity is still incompletely understood. In support of a protective role, HCV infectivity for chimpanzees has been neutralized by *in vitro* treatment with antibodies⁵, and infection outcome in humans was predicted by sequence changes in the hypervariable-1 region of the E2 envelope glycoprotein, a major target of the antibody response, that occurred simultaneously with antibody seroconversion⁶. These sequence changes are thought to represent escape mutations, and so

HCV almost certainly adapts to immune selection pressure exerted by antibodies and, as will be discussed below, CD8⁺ T lymphocytes. On the other hand, the role of naturally acquired antibodies in protection has been questioned because they do not prevent reinfection of immune chimpanzees or humans^{7,8}. Moreover, resolution of HCV infection can occur without the development of anti-HCV antibodies in chimpanzees⁹ and in the absence of seroconversion by standard assays in humans¹⁰.

Delineating the contribution of anti-HCV antibodies in infection outcome will probably depend on development of *in vitro* culture models for measuring their neutralizing capacity. Important progress towards this goal has recently been described. Infectious lentiviral pseudotype particles bearing native HCV envelope glycoproteins have been used to show cross-viral genotype neutralization of HCV by serum antibodies from chronically infected subjects^{11–13}. However, these antibodies are rare in individuals who resolve infection^{11–13}, suggesting that other mechanisms of adaptive immunity are important contributors to viral clearance.

T-cell immunity, HCV replication and liver damage

Acute infection

HCV genomes and proteins have been visualized within human and chimpanzee hepatocytes^{14,15}, but estimates of the proportion infected are variable and uncertain. Viraemia, or presence of HCV RNA genomes in the bloodstream, is therefore used as a surrogate for intra-hepatic HCV replication. Three broad patterns of acute phase replication were identified shortly after the discovery of HCV¹⁶ and are now being reinterpreted as T-cell responses to viral proteins are defined. HCV RNA genomes appear in the plasma within a few days of infection and typically peak 6–10 weeks later regardless of outcome^{16–18}. A pattern of poorly controlled viraemia predicts persistence (Fig. 1a) that may be explained, at least in part, by the failure of some individuals to generate detectable CD4⁺ and CD8⁺ T-cell responses^{9,19–22} (Fig. 1a). Further along the spectrum are individuals with transient (Fig. 1b) or permanent (Fig. 1c) control of viral replication that is temporally associated with late onset HCV-specific CD4⁺ and CD8⁺ T-cell responses^{9,19–22}. Why cellular immunity is often delayed for several weeks, even in humans and chimpanzees who ultimately clear the infection, is not yet understood^{9,22–24}. Viraemia also often rebounds after the infection is substantially controlled^{21–23} and can be a prelude to viral persistence.

¹Center for Vaccines and Immunity, Columbus Children's Research Institute, Columbus, Ohio OH 43205, USA. ²College of Medicine and Public Health, Ohio State University, Columbus, Ohio, USA.

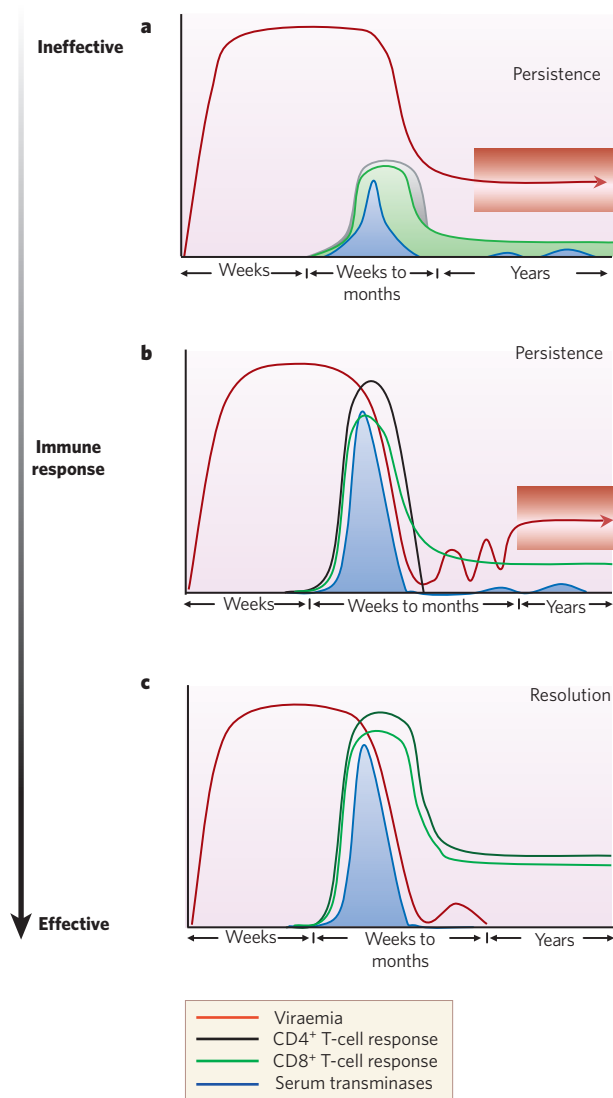


Figure 1 | Schematic representation of the cellular immune response during acute HCV infection. **a**, Viraemia (red line) is present early, and although it usually falls from peak levels, it is never controlled. Persistent infection ensues, with plasma viral levels that vary widely between individuals. CD4⁺ and CD8⁺ T-cell responses (black and green lines, respectively) and rises in serum transaminase (blue) in this setting remain poorly characterized, and may be weak or even absent. Progressive shading indicates the high degree of variability in responses between individuals. **b**, Viraemia may persist for many weeks in the absence of a demonstrable cellular immune response. The delayed onset of CD4⁺ and CD8⁺ T-cell responses is associated with transient control of viraemia and a variable rise in transaminases. However, following contraction of the CD4⁺ T-cell response, viraemia rebounds and ultimately infection persists. Detectable CD8⁺ T-cell responses may persist despite chronic viraemia. Shading indicates variability in viral load between individuals in the chronic phase. **c**, Although viraemia arises early and T-cell responses are delayed, the virus becomes undetectable in plasma following emergence of CD4⁺ and CD8⁺ T-cell responses, which are often coincident with a variable rise in serum transaminases. A rebound in viraemia may occur before final viral clearance.

Liver damage is monitored by levels of hepatic enzymes (transaminases) released into the serum following hepatocellular injury. Acute hepatitis C is probably immunopathological because it coincides temporally with expansion of virus-specific CD8⁺ T cells or cytotoxic T lymphocytes (CTLs)^{9,22–24} and their acquisition of an activated phenotype²⁵. Curiously, acute infection can sometimes resolve without a serum transaminase increase. This could reflect non-cytolytic control of some infections by T-cell-derived cytokines such as interferon- γ

(IFN- γ)^{23,26,27}, but it is also possible that relatively few hepatocytes are infected in these individuals, which would limit the extent of CTL-mediated liver damage.

Chronic infection

As discussed in detail below, HCV-specific CD8⁺ T cells can survive for years in the persistently infected liver and thus might at least partially control ongoing viral replication and/or contribute to progressive liver disease. With regard to viral replication, it is noteworthy that levels of viraemia remain relatively stable over time in subjects with chronic infections even though they can differ widely between individuals (Fig. 1). Data regarding the relationship between intrahepatic CTLs and viral load have been conflicting^{28–31}, and thus whether they contribute to control of HCV replication in chronic hepatitis C or explain the wide variation in viraemia between individuals is not known. Similarly, although virus-specific CTLs are concentrated within the liver in chronic infection^{29,30,32}, they have only inconsistently been correlated with disease severity^{29,31}. Some studies have documented an association between elevated transaminases and liver infiltration by CD8⁺ T cells^{31,33}, but it is possible that only a minority of the infiltrate is HCV specific³⁴. Intrahepatic pro-inflammatory cytokine messenger RNA (mRNA) levels have been correlated with severity of portal inflammation and liver fibrosis^{35,36}. However, it is unclear whether HCV-specific liver infiltrating lymphocytes are the major cellular source of pro-inflammatory cytokines, such as IFN- γ , in HCV-associated chronic hepatitis³⁶. Importantly, recruitment of other inflammatory cell types such as macrophages that can mediate tissue injury also occurs in chronic hepatitis C³⁶. Thus, studies to weigh the contribution of both antigen-specific and antigen-independent mechanisms in chronic hepatocellular injury are needed.

Parsing T-cell responses in resolving and persistent infections

Features that distinguish T-cell responses providing transient versus permanent viral control are not fully defined. Comparisons have focused on the number of epitopes targeted by acute-phase T cells, their frequency in blood and their fate as infections clear or persist. CD8⁺ T-cell responses are perhaps the best characterized. Successful responses generally target multiple major histocompatibility complex (MHC) class I-restricted epitopes in structural and non-structural HCV proteins^{9,24,37,38}, and frequencies against individual epitopes often exceed 3–4% (ref. 37). Interestingly, late expansion of HCV-specific CD8⁺ T cells may be accompanied by a further delay in their production of IFN- γ ^{22,24,37}, although the significance of this phenomenon to infection outcome is uncertain.

Infections that follow a chronic course are usually marked by low frequencies of CTLs targeting few epitopes^{28,39–42}, although occasionally the strength and breadth of the response can approximate those of resolving infections⁴³. HCV-specific CTLs are not necessarily deleted as infections persist and, although difficult to detect in blood, can survive in the persistently infected liver for years^{29,30,32,44,45} and in some cases target a surprisingly broad array of epitopes. For example, intrahepatic CTL populations recognizing eight HCV epitopes were consistently recovered from a chimpanzee throughout years of persistent infection^{45,46}, and CTL clones isolated at diverse timepoints often retained use of identically recombined T-cell receptor (TCR) α and β chains⁴⁷, providing genetic evidence of their remarkable stability despite ongoing viral replication. These intrahepatic populations are almost certainly a remnant of a once robust acute phase response.

Strong CD8⁺ T-cell immunity in acute resolving hepatitis C is matched by vigorous, sustained CD4⁺ T-cell proliferation to multiple recombinant structural and non-structural viral proteins^{19–21}. By contrast, HCV antigen-driven proliferation in individuals who develop persistent infections is usually weak or absent when compared with spontaneously resolving infections^{19–22}. Nevertheless, transient CD4⁺ T-cell proliferative responses that are indistinguishable from those in acute resolving infections have been described^{21,22}, and although possibly rare do indicate that HCV can evade what appears to be initially

robust CD4⁺ T helper (T_H) cell activity. Very importantly, permanent loss of this HCV-specific proliferation during acute hepatitis C predicts persistence^{20,21}.

In addition to blunted and transient proliferative responses, CD4⁺ T cells from persistently infected individuals target few MHC class II-restricted epitopes. For instance, in one study very few HCV-specific CD4⁺ T cell lines were established from humans with persistent infection, whereas those with resolved infections recognized up to 14 different HCV epitopes⁴⁸. Despite the broad diversity of epitopes targeted, preliminary studies in a chimpanzee with acutely resolving infection suggest that evolution of a multi-specific CD4⁺ T-cell response can be complex, initially focusing on a limited number of dominant epitopes and then spreading to additional targets only after viraemia is mostly controlled⁴⁹. Potential consequences for infection outcome of a T_H response that is transiently narrow during a critical period of viral control have not yet been fully explored.

Memory T cells and HCV resolution

Control of viraemia is associated with the contraction of detectable virus-specific CD4⁺ and CD8⁺ T-cell responses, almost certainly by programmed death as normally occurs at the end of cellular immune responses. This late phase of the response is often marked by resurgent HCV replication that may reflect a delicate balance between slow clearance of intrahepatic viral genomes and an easing of immune control as virus-specific T cells die (Fig. 1b, c). As noted above, this rebound in viraemia can lead to viral persistence after a transient period of control, and it may reflect active evasion of T-cell responses in the very late stages of acute hepatitis C. In contrast, infections that are successfully controlled result in durable memory populations. For instance, most subjects who resolved an accidental infection with a single source of HCV-contaminated immunoglobulin⁴⁰ had strong HCV-specific T-cell immunity in blood 18 years later, even though serologic responses to the virus had waned in the majority. Durable intrahepatic memory is probably also established, because T cells recognizing HCV antigens have been recovered from the livers of chimpanzees several years after spontaneous clearance of infection²⁴.

T-cell memory may explain a substantially lower rate of HCV persistence in re-exposed humans with a history of acute resolving hepatitis C⁵⁰. In support of this concept, immune chimpanzees have been shown to be susceptible to reinfection, but there were marked reductions in the duration and peak of viraemia^{24,51,52} coinciding temporally with massive CD4⁺ and CD8⁺ T-cell recall responses²⁴ (illustrated schematically in Fig. 2). Importantly, antibody-mediated depletion of CD8⁺ T cells from immune chimpanzees prolonged viraemia after rechallenge with the same HCV strain, and viral clearance was precisely correlated with recovery of HCV-specific CD8⁺ T cells within the liver²⁴. Treatment with anti-CD4 antibodies in a parallel experiment resulted in HCV persistence, revealing the importance of memory CD4⁺ T cells to infection outcome³⁸ (Fig. 2).

Mechanisms of HCV persistence

Descriptions of the repertoire, frequency and fate of CD4⁺ and CD8⁺ T cells in acute and chronic hepatitis C have provided limited insight into their subversion by HCV. A variety of postulated mechanisms are summarized in Fig. 3 and discussed below. It should be noted, however, that few enjoy substantial experimental support, and none can fully account for the inability to generate or sustain a CD4⁺ T_H cell response.

It is important to emphasize that, in the absence of advanced liver disease, defects in cellular immunity appear exquisitely HCV specific. Any proposed model of evasion should therefore explain an antigen-specific lesion in cellular immunity lasting for decades. This would seem to exclude, for example, direct cytopathic destruction of infected T lymphocytes or antigen-presenting cells (APCs). Indirect suppression of T-cell activity by viral proteins binding ubiquitous surface receptors is also not easily reconciled with apparently normal immune function in persistently infected individuals. Immunomodulatory activity of HCV pro-

teins like the viral nucleocapsid is nonetheless supported by *in vitro* cell-culture models^{53,54}, and it is possible that threshold protein concentrations required for suppression are reached only in the infected liver or perhaps during acute hepatitis C when viral replication peaks.

A number of mechanisms consistent with an HCV-specific defect in immunity have been proposed. Two that are conceptually attractive despite limited experimental support include an inability of effector T cells to move to the infected liver²³ and impaired antigen presentation. The potential for impaired antigen presentation in HCV infection merits attention because it could explain defects in immunity ranging from an apparent absence of HCV-specific T cells in some individuals to substantially delayed or nonsustained responses in others. Dendritic cells (DCs), the primary dedicated APC, may be particularly susceptible to a potentially potent and multi-faceted attack on innate immune mechanisms (see the review in this issue by Gale and Foy, page 939) essential for their maturation and effective antigen presentation to T cells. For instance, although natural killer (NK) cells are an important direct mediator of the innate immune response that have been implicated in the control of HCV infection⁵⁵ and may be inhibited by the HCV envelope E2 protein^{56,57}, they are also potent activators of DCs. However, recent data from an *in vitro* cell-culture model indicate that negative regulatory signals delivered to NK cells in the persistently HCV-infected liver could interrupt this ability to mediate DC maturation⁵⁸. Inhibition of antigen presentation in HCV infection might also be mediated more directly by the effects of viral proteins on DCs. Monocyte-derived DC from the blood of persistently infected individuals may display defects in maturation or stimulation of allogeneic T-cell responses^{59–61}. These have not been consistent findings in the literature, however^{62,63}, and permanent global defects in DC function extending into the chronic phase of infection are not clinically evident. On the other hand, a transient defect in DC function during acute hepatitis C could be sufficient to alter the timing or vigour of T-cell immunity and favour persistence.

Three other mechanisms of immune evasion that include mutational escape of epitopes, functional anergy and regulatory T-cell activity enjoy more substantial experimental support, even if their role in HCV persistence is not proven. Most of the published studies so far have focused on how these mechanisms alter HCV-specific CD8⁺ T cell immunity; their relevance to loss of CD4⁺ T cell help is still mostly unexplored.

Mutational escape of HCV epitopes

In common with other highly mutable RNA viruses, HCV is well adapted to generate genomic diversity that can be exploited to evade CD8⁺ T cells. The viral NS5b protein, an error-prone RNA-dependent RNA polymerase, could be considered an immune evasion molecule because it prolifically generates minor viral variants⁶⁴ with the potential to evade recognition. Epitopes may be lost because amino-acid substitutions result in proteasomal destruction or impaired binding to MHC molecules⁶⁵. Alternatively, amino-acid changes can alter CTL recognition of variant peptide–MHC complexes⁶⁵.

One key criticism of the mutational escape hypothesis is that, even with a high mutation rate, HCV is unlikely to evade multi-specific CTL responses found in some individuals who develop chronic infection. Delaying expansion and acquisition of effector functions by CD8⁺ T cells for weeks during acute infection could favour accumulation of variants with the potential to evade immune control. This delay on its own is probably not sufficient to facilitate escape from a multi-specific CTL response. When combined with the spontaneous loss of CD4⁺ T-cell help that precedes HCV persistence, however, CD8⁺ T-cell function might be impaired to the point where they are no longer capable of mediating viral clearance, but still exert sufficient immune pressure to select for variant viruses. The concept that lack of help undermines the effectiveness of HCV-specific CD8⁺ T cells is supported by recent data from immune chimpanzees³⁸, where re-challenge with the same HCV strain following antibody-mediated depletion of CD4⁺ T cells resulted in persistent infections associated with CTL escape mutations in multiple MHC class I-restricted epitopes³⁸ (Fig. 2). Other immuno-

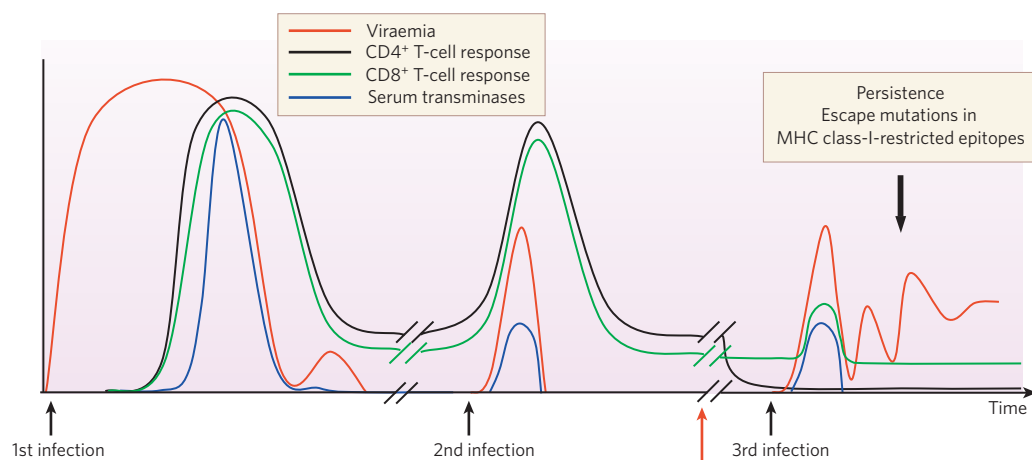


Figure 2 | Schematic representation of memory response to homologous HCV re-challenge, and outcome of a third challenge following CD4⁺ T-cell depletion. Following initial rechallenge, the onset of viraemia (red line) is associated with relatively rapid onset of strong CD4⁺ and CD8⁺ memory T-cell responses (black and green lines, respectively). The level of viraemia is lower than in primary infection, and viral clearance is more rapid. Associated transaminase rises (blue line) are also reduced. Following CD4⁺ T-cell depletion (indicated by a red arrow), viral rechallenge is associated with a weaker CD8⁺ T-cell response, which partly controls viraemia but is followed by viral rebound and persistence as escape mutations emerge in MHC class-I-restricted epitopes.

logical factors such as the TCR diversity used in HCV-specific CD8⁺ T-cell responses may also be a factor in the development of escape mutations⁴⁷, with narrower responses less able to constrain the emergence of viral variants.

CD8⁺ T-cell mediated selection pressure against HCV was first demonstrated in the chimpanzee model⁴⁵. Evidence for escape mutations in human HCV infections has accumulated slowly, especially as the sequence of the transmitted virus is rarely known. Early indirect evidence for CTL escape mutations was inferred by comparison of circulating sequences in chronically HCV-infected individuals with prototypical epitopes⁶⁶. These data have recently been reinforced by studies where viral sequences were available either from the source or during early infection, in which the development of persistent viraemia was associated with escape mutations in targeted MHC class-I-restricted epitopes^{67–69}. Recent population-based approaches have also provided further support for CTL-driven HCV evolution^{67,70}.

Studies in human (HIV) and simian (SIV) immunodeficiency virus infections indicate that escape mutations may exact and be constrained by a 'fitness cost' to replication that varies between epitopes. Following transmission of HIV or SIV to hosts expressing non-selecting MHC, certain epitopes with potentially high associated fitness cost revert to wild-type sequence^{71,72}. Others, presumably associated with low replicative impairment, revert slowly or not at all^{71,73}. Recent analyses of HCV-infected cohorts have demonstrated that mutations tending toward viral consensus do occur in prototypical epitopes unrestricted by host MHC alleles^{67,69,70}, which may indicate reversion of virus to more replicatively fit ancestral sequences. However, further studies are required to define the extent to which fitness cost constrains the development of escape mutations within HCV epitopes.

It is important to note that the phenomenon of CTL escape mutation may influence not just viral quasispecies within individuals, but also the dominant viral species circulating in human populations. Escape mutations within HIV epitopes restricted by common MHC class I alleles may impart a 'footprint' of associated polymorphisms upon viral species within a population⁷⁴. Furthermore, CTL escape mutations in epitopes with negligible fitness cost may persist in HIV genomes transmitted to individuals without the restricting MHC class I allele, to the point of becoming the dominant sequence observed at a population level⁷³. Thus, immunodominant HIV epitopes restricted by MHC alleles that are common within a population may be lost from circulating viral species⁷⁵. Preliminary evidence indicates that commonly expressed HLA alleles may also influence HCV sequences at a population level (S. Gaudieri, personal communication).

The relationship between fitness cost associated with escape mutation and clearance of HCV infection remains largely unexplored. It is

conceivable that CD8⁺ T cells mediating viral clearance target a set of protective epitopes functionally constrained from mutation. However, there is as yet no experimental support for this hypothesis in an infection where patterns of epitope dominance are usually neither obvious nor easily predicted from HLA class I haplotypes of infected subjects⁷⁶.

The hypothesis that CD4⁺ T cells can also exert immune selection pressure is attractive because, as noted above, the response might initially focus on a limited set of dominant epitopes and may be influenced by MHC class II allelic associations with infection outcome³. However, although amino-acid substitutions in HLA class II epitopes of HCV can skew patterns of cytokine expression⁷⁷ and antagonize⁷⁸ or abrogate⁷⁹ T_H cell activity, few MHC class II epitopes have been studied so far and despite persistent HCV replication, amino-acid substitutions are rarely observed⁸⁰. Perhaps most importantly, formal statistical proof that the rate of mutations that result in amino-acid substitutions is increased in MHC class II-restricted versus unrestricted epitopes or flanking regions of the HCV genome is still lacking.

It should be noted that many T-cell epitopes are intact in persistently replicating HCV genomes and do not mutate despite an intense and focal CD8⁺ T-cell response⁴³. These observations suggest that mutational escape is not the only mechanism for evading antigen-specific T cells. Experimental evidence for two additional mechanisms, functional anergy and generation of regulatory T-cell populations, indicate that there are multiple complementary pathways to establish or maintain viral persistence in the presence of HCV-specific T cells.

Deletion or anergy

Although deletion of antigen-specific CD8⁺ T cells has been demonstrated in chronic murine lymphocytic choriomeningitis virus (LCMV) infection⁸¹, no data are yet available as to whether such mechanisms affect HCV-specific CD8⁺ T cells. However, recent studies of HCV-specific CD8⁺ T cells have indicated that these cells may be functionally impaired, or anergic, in chronic disease^{41,82,83} and, consistent with this loss of function, may exhibit phenotypic alterations characteristic of early stages of differentiation⁸⁴. Interestingly, congruent with chronic HCV infection, in LCMV infection CD8⁺ T-cell anergy has been associated with failing or absent virus-specific CD4⁺ T-cell responses^{85,86}.

Although these findings suggest a possible role for CD8⁺ T-cell anergy in HCV persistence, their significance remains uncertain as virus-specific CD8⁺ T cells were also impaired after resolution of infection in at least one of these studies⁸². Furthermore, phenotypic alterations of CMV-specific CD8⁺ T cells have also been described in chronic HCV infection despite a lack of clinically evident dysfunction⁸⁷. Moreover, the majority of phenotypic studies of HCV-specific

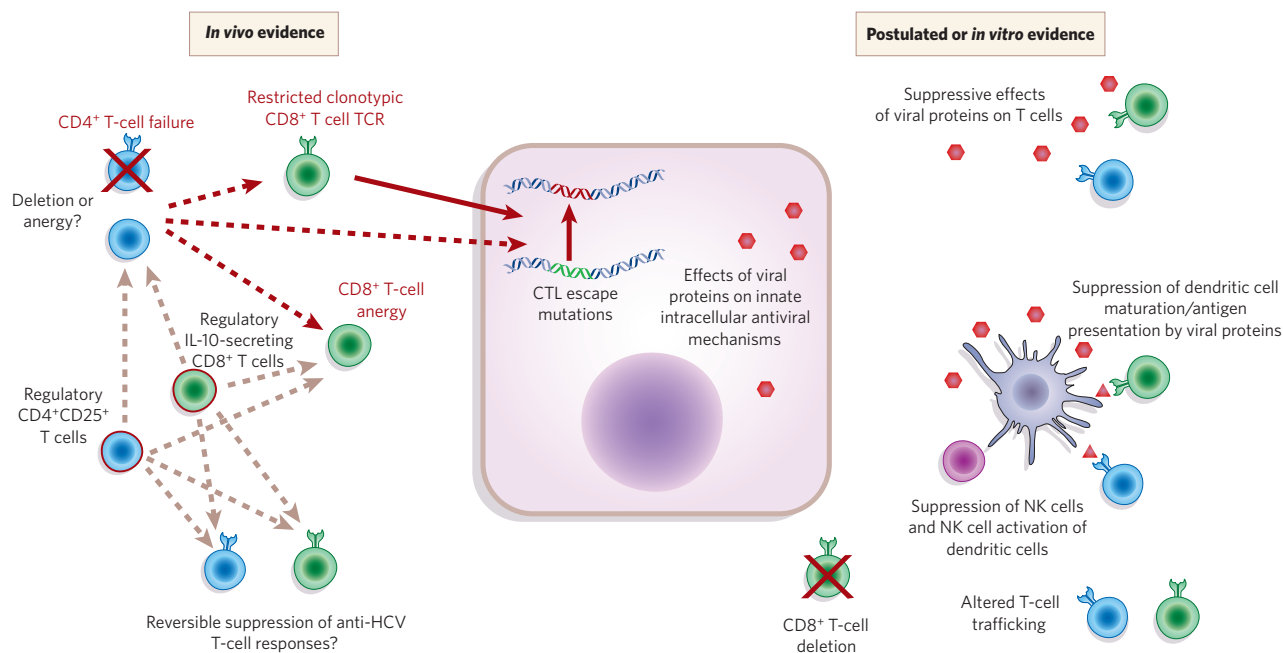


Figure 3 | Possible mechanisms of immune evasion by the hepatitis C virus. Red lines and headings represent mechanisms for which there is the most supporting *in vivo* experimental data; dotted lines indicate that although the mechanisms indicated may be involved in viral persistence, their pathways

remain unclear. Large crosses indicate deletion of cells, whereas triangles indicate inhibition of antigen presentation. Interactions of viral products with intracellular pathways and the consequent effects on innate immunity are covered in the accompanying review by Gale and Foy (page 939).

CD8⁺ T cells have been performed on peripheral blood; it remains possible that the characteristics of virus-specific CTLs may differ within the liver, the primary site of infection. Thus, further examination of HCV-specific T cell phenotype and function is required before definitive conclusions can be drawn.

Although persisting CTL responses are detectable in some chronically HCV-infected individuals, virus-specific CD4⁺ T cell responses are generally weak or absent when assessed using functional methods of identification^{20,40,48,88}. However, it remains unclear whether HCV-specific CD4⁺ T cells are present but functionally impaired or have been lost from the repertoire. In a recent study employing MHC class II tetramers to identify antigen-specific cells independent of function, HCV-specific CD4⁺ T cells were not detected in chronically viraemic individuals⁸⁹. By contrast, a study in which upregulation of the activation marker CD25 in response to antigenic stimulation was used to identify HCV-specific CD4⁺ T cells indicated that such cells might be present, but functionally altered⁹⁰. It is also possible that loss of virus-specific CD4⁺ T cells may be somewhat epitope dependent, with cells specific for the core antigen being more readily detected than those directed against the non-structural proteins^{91,92}. Interestingly, a sub-population of CD4⁺ T cells specific for HCV core that persist in chronic infection may be characterized by altered patterns of cytokine production including secretion of the anti-inflammatory cytokine interleukin (IL)-10 (ref. 91) or impaired secretion of the survival cytokine IL-2 (ref. 92).

Regulatory T-cell populations

A subset of HCV-specific CD8⁺ T-cell lines derived from the liver of a persistently infected subject was found to produce the immune suppressive cytokine IL-10, providing the first suggestion of MHC class-I-restricted antigen-specific regulatory activity with the potential to suppress antiviral T cells⁹³. These data were reinforced by more recent observations that intrahepatic CD8⁺ T cells from persistently infected subjects suppressed the *in vitro* proliferative responses of liver-derived lymphocytes in an HCV-specific and IL-10-dependent manner⁹⁴.

In addition to the presence of possible intrahepatic regulatory CD8⁺ T-cell populations, the proportion of CD4⁺CD25⁺ regulatory T cells may be elevated in the peripheral circulation of some chronically

HCV-infected subjects when compared with recovered or uninfected individuals^{95–97}. Depletion of CD4⁺CD25⁺ T cells from peripheral blood increased the frequency of functional HCV-specific CD8⁺ T cells in *in vitro* assays^{95–98}, indicating the potential for regulatory activity to suppress antiviral responses. Interpretation of this observation is not straightforward, however, because suppression extended to CD8⁺ T cells targeting cytomegalovirus, Epstein–Barr virus and influenza virus^{97,98}. Our understanding of regulatory CD4⁺ T cells in viral persistence is further complicated by uncertainty over their localization to secondary lymphoid organs or liver where antigen-specific interactions relevant to HCV persistence are more likely to occur.

In summary, HCV persistence is predicted by a failure to generate or sustain CD4⁺ T-cell responses, and this outcome can be recapitulated by anti-CD4 antibody treatment of immune chimpanzees, suggesting that an HCV-specific loss of T-cell help is a central event required for immune evasion. Although lacking in critical detail and experimental proof, one working model of HCV persistence is that the virus sets in train a series of cascading events where interference with innate immunity causes a defect in CD4⁺ T-cell help. Depending on the severity of this impairment, CD8⁺ T cells essential for viral clearance either succeed or fail. Absence of adequate help could provide a common explanation for two of the major defects observed in the CD8⁺ T-cell compartment, specifically mutational escape of MHC class-I-restricted epitopes and functional anergy. It is, however, too soon to discount the importance of other more direct mechanisms of CD8⁺ T cell inactivation, particularly in an organ like the liver with its own unique populations of APCs that can modulate immunity⁹⁹.

Vaccines for HCV: opportunities and risks

The temporal kinetic relationship between the onset of T-cell responses and control of viraemia, and the altered outcome of infection in immune chimpanzees depleted of CD4⁺ and CD8⁺ subsets, clearly indicate that cellular immunity is critical in prevention of HCV persistence. These findings, combined with the observation that immunity can even provide protection against different HCV genotypes¹⁰⁰, have raised hopes for a safe and effective vaccine (see the review in this issue by Houghton and Abrignani, page 961). Despite

these positive developments, from the point of view of adaptive immunity to the virus, it may be too soon to predict success. Our understanding of humoral immunity to HCV, particularly the potential for antibody cross-reactivity among genetically diverse isolates and the role of neutralizing antibodies, is still rudimentary. With regard to cellular immunity, we still lack knowledge of how HCV is cleared and it is not at all certain that robust virus-specific responses as assessed by currently employed assays, in particular production of IFN- γ , are adequate surrogates for protective T-cell immunity. Perhaps the greatest risk comes from our poor insight into how HCV inactivates primed CD4⁺ T_H cells even after they have contributed to transient control of viraemia. Our ability to safely harness protective immunity by vaccination may depend on the answers to these very basic questions. ■

- Bowmen, D. G. & Walker, C. M. The origin of quasispecies: cause or consequence of chronic hepatitis C viral infection? *J. Hepatol.* **42**, 408–417 (2005).
- Pearlman, B. L. Hepatitis C treatment update. *Am. J. Med.* **117**, 344–352 (2004).
- Shoukry, N. H., Cawthon, A. G. & Walker, C. M. Cell-mediated immunity and the outcome of hepatitis C virus infection. *Annu. Rev. Microbiol.* **58**, 391–424 (2004).
- Pawlotsky, J. M. Diagnostic tests for hepatitis C. *J. Hepatol.* **31** (Suppl 1), 71–79 (1999).
- Farci, P. et al. Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated *in vitro* neutralization. *Proc. Natl Acad. Sci. USA* **91**, 7792–7796 (1994).
- Farci, P. et al. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339–344 (2000).
- Farci, P. et al. Lack of protective immunity against reinfection with hepatitis C virus. *Science* **258**, 135–140 (1992).
- Lai, M. E. et al. Hepatitis C virus in multiple episodes of acute hepatitis in polytransfused thalassaemic children. *Lancet* **343**, 388–390 (1994).
- Cooper, S. et al. Analysis of a successful immune response against hepatitis C virus. *Immunity* **10**, 439–449 (1999).
- Post, J. J. et al. Clearance of hepatitis C viremia associated with cellular immunity in the absence of seroconversion in the hepatitis C incidence and transmission in prisons study cohort. *J. Infect. Dis.* **189**, 1846–1855 (2004).
- Bartosch, B. et al. *In vitro* assay for neutralizing antibody to hepatitis C virus: evidence for broadly conserved neutralization epitopes. *Proc. Natl Acad. Sci. USA* **100**, 14199–14204 (2003).
- Logvinoff, C. et al. Neutralizing antibody response during acute and chronic hepatitis C virus infection. *Proc. Natl Acad. Sci. USA* **101**, 10149–10154 (2004).
- Meunier, J. C. et al. Evidence for cross-genotype neutralization of hepatitis C virus pseudo-particles and enhancement of infectivity by apolipoprotein C1. *Proc. Natl Acad. Sci. USA* **102**, 4560–4565 (2005).
- Negro, F. et al. Detection of intrahepatic replication of hepatitis C virus RNA by *in situ* hybridization and comparison with histopathology. *Proc. Natl Acad. Sci. USA* **89**, 2247–2251 (1992).
- Nouri-Aria, K. T., Sallie, R., Mizokami, M., Portmann, B. C. & Williams, R. Intrahepatic expression of hepatitis C virus antigens in chronic liver disease. *J. Pathol.* **175**, 77–83 (1995).
- Abe, K., Inchauspe, G., Shikata, T. & Prince, A. M. Three different patterns of hepatitis C virus infection in chimpanzees. *Hepatology* **15**, 690–695 (1992).
- Beach, M. J. et al. Temporal relationships of hepatitis C virus RNA and antibody responses following experimental infection of chimpanzees. *J. Med. Virol.* **36**, 226–237 (1992).
- Alter, H. J. et al. Evaluation of branched DNA signal amplification for the detection of hepatitis C virus RNA. *J. Viral Hepatol.* **2**, 121–132 (1995).
- Missale, G. et al. Different clinical behaviors of acute hepatitis C virus infection are associated with different vigor of the anti-viral cell-mediated immune response. *J. Clin. Invest.* **98**, 706–714 (1996).
- Diepolder, H. M. et al. Possible mechanism involving T-lymphocyte response to non-structural protein 3 in viral clearance in acute hepatitis C virus infection. *Lancet* **346**, 1006–1007 (1995).
- Gerlach, J. T. et al. Recurrence of hepatitis C virus after loss of virus-specific CD4⁺ T-cell response in acute hepatitis C. *Gastroenterology* **117**, 933–941 (1999).
- Thimme, R. et al. Determinants of viral clearance and persistence during acute hepatitis C virus infection. *J. Exp. Med.* **194**, 1395–1406 (2001).
- Thimme, R. et al. Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease. *Proc. Natl Acad. Sci. USA* **99**, 15661–15668 (2002).
- Shoukry, N. H. et al. Memory CD8⁺ T cells are required for protection from persistent hepatitis C virus infection. *J. Exp. Med.* **197**, 1645–1655 (2003).
- Lechner, F. et al. CD8⁺ T lymphocyte responses are induced during acute hepatitis C virus infection but are not sustained. *Eur. J. Immunol.* **30**, 2479–2487 (2000).
- Frese, M. et al. Interferon-gamma inhibits replication of subgenomic and genomic hepatitis C virus RNAs. *Hepatology* **35**, 694–703 (2002).
- Lanford, R. E. et al. Antiviral effect and virus-host interactions in response to alpha interferon, gamma interferon, poly(I)-poly(C), tumor necrosis factor alpha, and ribavirin in hepatitis C virus subgenomic replicons. *J. Virol.* **77**, 1092–1104 (2003).
- Rehermann, B. et al. Quantitative analysis of the peripheral blood cytotoxic T lymphocyte response in patients with chronic hepatitis C virus infection. *J. Clin. Invest.* **98**, 1432–1440 (1996).
- Nelson, D. R. et al. The role of hepatitis C virus-specific cytotoxic T lymphocytes in chronic hepatitis C. *J. Immunol.* **158**, 1473–1481 (1997).
- Wong, D. K. et al. Liver-derived CTL in hepatitis C virus infection: breadth and specificity of responses in a cohort of persons with chronic infection. *J. Immunol.* **160**, 1479–1488 (1998).
- Freeman, A. J. et al. The presence of an intrahepatic cytotoxic T lymphocyte response is associated with low viral load in patients with chronic hepatitis C virus infection. *J. Hepatol.* **38**, 349–356 (2003).
- He, X. S. et al. Quantitative analysis of hepatitis C virus-specific CD8⁺ T cells in peripheral blood and liver using peptide-MHC tetramers. *Proc. Natl Acad. Sci. USA* **96**, 5692–5697 (1999).
- Kurokohchi, K. et al. CD28-negative CD8-positive cytotoxic T lymphocytes mediate hepatocellular damage in hepatitis C virus infection. *J. Clin. Immunol.* **23**, 518–527 (2003).
- Abbrignani, S. Bystander activation by cytokines of intrahepatic T cells in chronic viral hepatitis. *Semin. Liver Dis.* **17**, 319–322 (1997).
- Napoli, J., Bishop, G. A., McGuinness, P. H., Painter, D. M. & McCaughan, G. W. Progressive liver injury in chronic hepatitis C infection correlates with increased intrahepatic expression of Th1-associated cytokines. *Hepatology* **24**, 759–765 (1996).
- McGuinness, P. H., Painter, D., Davies, S. & McCaughan, G. W. Increases in intrahepatic CD68 positive cells, MAC387 positive cells, and proinflammatory cytokines (particularly interleukin 18) in chronic hepatitis C infection. *Gut* **46**, 260–269 (2000).
- Lechner, F. et al. Analysis of successful immune responses in persons infected with hepatitis C virus. *J. Exp. Med.* **191**, 1499–1512 (2000).
- Grakoui, A. et al. HCV persistence and immune evasion in the absence of memory T cell help. *Science* **302**, 659–662 (2003).
- Koziel, M. J. et al. Hepatitis C virus (HCV)-specific cytotoxic T lymphocytes recognize epitopes in the core and envelope proteins of HCV. *J. Virol.* **67**, 7522–7532 (1993).
- Takaki, A. et al. Cellular immune responses persist and humoral responses decrease two decades after recovery from a single-source outbreak of hepatitis C. *Nature Med.* **6**, 578–582 (2000).
- Wedemeyer, H. et al. Impaired effector function of hepatitis C virus-specific CD8⁺ T cells in chronic hepatitis C virus infection. *J. Immunol.* **169**, 3447–3458 (2002).
- Lauer, G. M. et al. High resolution analysis of cellular immune responses in resolved and persistent hepatitis C virus infection. *Gastroenterology* **127**, 924–936 (2004).
- Urbani, S. et al. Heterologous T cell immunity in severe hepatitis C virus infection. *J. Exp. Med.* **201**, 675–680 (2005).
- Koziel, M. J. et al. Intrahepatic cytotoxic T lymphocytes specific for hepatitis C virus in persons with chronic hepatitis. *J. Immunol.* **149**, 3339–3344 (1992).
- Erickson, A. L. et al. The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes. *Immunity* **15**, 883–895 (2001).
- Kimura, Y., Gushima, T., Rawale, S., Kaumaya, P. & Walker, C. M. Escape mutations alter proteasome processing of major histocompatibility complex class I-restricted epitopes in persistent hepatitis C virus infection. *J. Virol.* **79**, 4870–4876 (2005).
- Meyer-Olson, D. et al. Limited T cell receptor diversity of HCV-specific T cell responses is associated with CTL escape. *J. Exp. Med.* **200**, 307–319 (2004).
- Day, C. L. et al. Broad specificity of virus-specific CD4⁺ T-helper-cell responses in resolved hepatitis C virus infection. *J. Virol.* **76**, 12584–12595 (2002).
- Shoukry, N. H., Sidney, J., Sette, A. & Walker, C. M. Conserved hierarchy of helper T cell responses in a chimpanzee during primary and secondary hepatitis C virus infections. *J. Immunol.* **172**, 483–492 (2004).
- Mehta, S. H. et al. Protection against persistence of hepatitis C. *Lancet* **359**, 1478–1483 (2002).
- Bassett, S. E. et al. Protective immune response to hepatitis C virus in chimpanzees rechallenged following clearance of primary infection. *Hepatology* **33**, 1479–1487 (2001).
- Major, M. E. et al. Previously infected and recovered chimpanzees exhibit rapid responses that control hepatitis C virus replication upon rechallenge. *J. Virol.* **76**, 6586–6595 (2002).
- Hahn, Y. S. Subversion of immune responses by hepatitis C virus: immunomodulatory strategies beyond evasion? *Curr. Opin. Immunol.* **15**, 443–449 (2003).
- Accapezzato, D. et al. Subversion of effector CD8⁺ T cell differentiation in acute hepatitis C virus infection: the role of the virus. *Eur. J. Immunol.* **34**, 438–446 (2004).
- Khakoo, S. I. et al. HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* **305**, 872–874 (2004).
- Crotta, S. et al. Inhibition of natural killer cells through engagement of CD81 by the major hepatitis C virus envelope protein. *J. Exp. Med.* **195**, 35–41 (2002).
- Tseng, C. T. & Klimpel, G. R. Binding of the hepatitis C virus envelope protein E2 to CD81 inhibits natural killer cell functions. *J. Exp. Med.* **195**, 43–49 (2002).
- Junishi, M. et al. Negative regulation of NK cell activities by inhibitory receptor CD94/NKG2A leads to altered NK cell-induced modulation of dendritic cell functions in chronic hepatitis C virus infection. *J. Immunol.* **173**, 6072–6081 (2004).
- Kanto, T. et al. Impaired allostimulatory capacity of peripheral blood dendritic cells recovered from hepatitis C virus-infected individuals. *J. Immunol.* **162**, 5584–5591 (1999).
- Auffermann-Gretzinger, S., Keffe, E. B. & Levy, S. Impaired dendritic cell maturation in patients with chronic, but not resolved, hepatitis C virus infection. *Blood* **97**, 3171–3176 (2001).
- Bain, C. et al. Impaired allostimulatory function of dendritic cells in chronic hepatitis C infection. *Gastroenterology* **120**, 512–524 (2001).
- Larsson, M. et al. Lack of phenotypic and functional impairment in dendritic cells from chimpanzees chronically infected with hepatitis C virus. *J. Virol.* **78**, 6151–6161 (2004).
- Piccoli, D. et al. Comparable functions of plasmacytoid and monocyte-derived dendritic cells in chronic hepatitis C patients and healthy donors. *J. Hepatol.* **42**, 61–67 (2005).
- Bukh, J., Miller, R. H. & Purcell, R. H. Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. *Semin. Liver Dis.* **15**, 41–63 (1995).
- Bowen, D. G. & Walker, C. M. Mutational escape from CD8⁺ T cell immunity: HCV evolution, from chimpanzees to man. *J. Exp. Med.* **201**, 1709–1714 (2005).
- Chang, K. M. et al. Immunological significance of cytotoxic T lymphocyte epitope variants in patients chronically infected by the hepatitis C virus. *J. Clin. Invest.* **100**, 2376–2385 (1997).
- Timm, J. et al. CD8 epitope escape and reversion in acute HCV infection. *J. Exp. Med.* **200**, 1593–1604 (2004).
- Tester, I. et al. Immune evasion versus recovery after acute hepatitis C virus infection from a shared source. *J. Exp. Med.* **201**, 1725–1731 (2005).
- Cox, A. L. et al. Cellular immune selection with hepatitis C virus persistence in humans. *J. Exp. Med.* **201**, 1741–1752 (2005).
- Ray, S. C. et al. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. *J. Exp. Med.* **201**, 1753–1759 (2005).

71. Leslie, A. J. *et al.* HIV evolution: CTL escape mutation and reversion after transmission. *Nature Med.* **10**, 282–289 (2004).
72. Friedrich, T. C. *et al.* Reversion of CTL escape-variant immunodeficiency viruses *in vivo*. *Nature Med.* **10**, 275–281 (2004).
73. Leslie, A. *et al.* Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *J. Exp. Med.* **201**, 891–902 (2005).
74. Moore, C. B. *et al.* Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**, 1439–1443 (2002).
75. Altfeld, M. *et al.* The majority of currently circulating human immunodeficiency virus type 1 clade B viruses fail to prime cytotoxic T-lymphocyte responses against an otherwise immunodominant HLA-A2-restricted epitope: implications for vaccine design. *J. Virol.* **79**, 5000–5005 (2005).
76. Lauer, G. M. *et al.* Comprehensive analysis of CD8⁺-T-cell responses against hepatitis C virus reveals multiple unpredicted specificities. *J. Virol.* **76**, 6104–6113 (2002).
77. Wang, J. H., Layden, T. J. & Eckels, D. D. Modulation of the peripheral T-cell response by CD4 mutants of hepatitis C virus: transition from a Th1 to a Th2 response. *Hum. Immunol.* **64**, 662–673 (2003).
78. Frasca, L. *et al.* Hypervariable region 1 variants act as TCR antagonists for hepatitis C virus-specific CD4⁺ T cells. *J. Immunol.* **163**, 650–658 (1999).
79. Wang, H. & Eckels, D. D. Mutations in immunodominant T cell epitopes derived from the nonstructural 3 protein of hepatitis C virus have the potential for generating escape variants that may have important consequences for T cell recognition. *J. Immunol.* **162**, 4177–4183 (1999).
80. Penna, A. *et al.* Intrahepatic and circulating HLA class II-restricted, hepatitis C virus-specific T cells: functional characterization in patients with chronic hepatitis C. *Hepatology* **35**, 1225–1236 (2002).
81. Moskopidis, D., Lechner, F., Pircher, H. & Zinkernagel, R. M. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. *Nature* **362**, 758–761 (1993).
82. Gruener, N. H. *et al.* Sustained dysfunction of antiviral CD8⁺ T lymphocytes after infection with hepatitis C virus. *J. Virol.* **75**, 5550–5558 (2001).
83. Urbani, S. *et al.* Virus-specific CD8⁺ lymphocytes share the same effector-memory phenotype but exhibit functional differences in acute hepatitis B and C. *J. Virol.* **76**, 12423–12434 (2002).
84. Appay, V. *et al.* Memory CD8⁺ T cells vary in differentiation phenotype in different persistent virus infections. *Nature Med.* **8**, 379–385 (2002).
85. Zajac, A. J. *et al.* Viral immune evasion due to persistence of activated T cells without effector function. *J. Exp. Med.* **188**, 2205–2213 (1998).
86. Khanolkar, A., Fuller, M. J. & Zajac, A. J. CD4 T cell-dependent CD8 T cell maturation. *J. Immunol.* **172**, 2834–2844 (2004).
87. Lucas, M. *et al.* Pervasive influence of hepatitis C virus on the phenotype of antiviral CD8⁺ T cells. *J. Immunol.* **172**, 1744–1753 (2004).
88. Rosen, H. R. *et al.* Frequencies of HCV-specific effector CD4⁺ T cells by flow cytometry: correlation with clinical disease stages. *Hepatology* **35**, 190–198 (2002).
89. Day, C. L. *et al.* *Ex vivo* analysis of human memory CD4 T cells specific for hepatitis C virus using MHC class II tetramers. *J. Clin. Invest.* **112**, 831–842 (2003).
90. Ulsenheimer, A. *et al.* Detection of functionally altered hepatitis C virus-specific CD4 T cells in acute and chronic hepatitis C. *Hepatology* **37**, 1189–1198 (2003).
91. MacDonald, A. J. *et al.* CD4 T helper type 1 and regulatory T cells induced against the same epitopes on the core protein in hepatitis C virus-infected persons. *J. Infect. Dis.* **185**, 720–727 (2002).
92. Semmo, N. *et al.* Preferential loss of IL-2-secreting CD4⁺ T helper cells in chronic HCV infection. *Hepatology* **41**, 1019–1028 (2005).
93. Koziel, M. J. *et al.* HLA class I-restricted cytotoxic T lymphocytes specific for hepatitis C virus. Identification of multiple epitopes and characterization of patterns of cytokine release. *J. Clin. Invest.* **96**, 2311–2321 (1995).
94. Accapezzato, D. *et al.* Hepatic expansion of a virus-specific regulatory CD8⁺ T cell population in chronic hepatitis C virus infection. *J. Clin. Invest.* **113**, 963–972 (2004).
95. Sugimoto, K. *et al.* Suppression of HCV-specific T cells without differential hierarchy demonstrated *ex vivo* in persistent HCV infection. *Hepatology* **38**, 1437–1448 (2003).
96. Cabrera, R. *et al.* An immunomodulatory role for CD4⁺CD25⁺ regulatory T lymphocytes in hepatitis C virus infection. *Hepatology* **40**, 1062–1071 (2004).
97. Boettler, T. *et al.* T cells with a CD4⁺CD25⁺ regulatory phenotype suppress *in vitro* proliferation of virus-specific CD8⁺ T cells during chronic hepatitis C virus infection. *J. Virol.* **79**, 7860–7867 (2005).
98. Rushbrook, S. M. *et al.* Regulatory T cells suppress *in vitro* proliferation of virus-specific CD8⁺ T cells during persistent hepatitis C virus infection. *J. Virol.* **79**, 7852–7859 (2005).
99. Bertolino, P., McCaughan, G. W. & Bowen, D. G. Role of primary intrahepatic T-cell activation in the 'liver tolerance effect'. *Immunol. Cell Biol.* **80**, 84–92 (2002).
100. Lanford, R. E. *et al.* Cross-genotype immunity to hepatitis C virus. *J. Virol.* **78**, 1575–1581 (2004).

Acknowledgements Our research was supported by Public Health Service grants to C.M.W. D.G.B. was supported by a C.J. Martin Fellowship from the National Health and Medical Research Council of Australia and an AstraZeneca Fellowship in Medical Research from the Royal Australasian College of Physicians.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.M.W. (walker@pediatrics.ohio-state.edu).

Challenges and successes in developing new therapies for hepatitis C

Raffaele De Francesco¹ and Giovanni Migliaccio¹

Hepatitis C virus (HCV) will continue to be a serious global health threat for many years to come because of the chronic nature of the infection, its high prevalence and the significant morbidity of the resulting disease. Recently, a small number of molecules have produced encouraging results in proof-of-concept clinical trials. At the same time, preclinical evidence is accumulating that development of resistance will eventually limit the efficacy of new drugs. Thus, combinations of multiple agents will be required to treat chronic HCV infection.

With an estimated 170 million infected individuals worldwide, hepatitis C virus (HCV) exacts a heavy toll on public health¹. Despite considerable reduction of the incidence of new infections, the prevalence of HCV infection is predicted to remain constant in the near future². Our current interferon (IFN)-based therapies are effective in only a fraction of the patients and are plagued with adverse effects (see the review by Hoofnagle in this issue, page 967). New treatment regimens are needed that are more efficacious and better tolerated by all patients. Investigators have taken several different approaches to address this pressing medical need. Major research efforts have focused on the identification of agents that inhibit specific steps in the life cycle of the virus. These 'HCV-targeted drugs' include small-molecule, orally bioavailable inhibitors of the HCV enzymes as well as nucleic-acid-based agents that attack the viral RNA. In addition, agents that can modulate the host immune response are being investigated for their ability to control and possibly eradicate HCV infection. In spite of the difficulties posed by the lack of readily available laboratory models of viral infection, a handful of investigational compounds have just started to show promising results in early-phase clinical trials (Table 1).

Drug-resistant viruses emerge rapidly under the selective pressures exerted by antiviral drugs, however. This is a major concern for successful anti-HCV therapy. The fast turnover rate and the intrinsic low fidelity of the HCV replication machinery endows the virus with the ability to fully explore its genome space and quickly come up with mutations that render it resistant to antiviral drugs. Each newly generated HCV genome is expected to exhibit, on average, one nucleotide change per replication cycle. As a consequence, even in untreated individuals, HCV exists as a genetically heterogeneous viral population, termed 'quasispecies'. Thus, the clinical success of HCV-targeted drugs will depend on their ability to suppress all viral variants as well as prevent the emergence of resistant viruses. In view of these considerations, it is crucial that the optimization of anti-HCV drug candidates is guided by the study of their spectrum of action on the different genotypes as well as by their resistance profile.

In the absence of an efficient *in vitro* infection system, preclinical studies of HCV antiviral resistance were made possible by the availability of HCV replicons (see Box 1 and the review in this issue by Lindenbach and Rice, page 933). So far, it has been possible to select replicons that are resistant to several HCV enzyme inhibitors as well as replicons that escape the action of small interfering RNAs (siRNAs)

directed against the HCV genome³. For all agents analysed so far, a single mutation in the target gene seemed sufficient for conferring drug resistance *in vitro*. Depending on the genetic make-up of the replicon, some of these resistance mutations affected the replication fitness in tissue culture. Given the many unknowns and the different requirements for efficient HCV replication *in vitro* and *in vivo* (see the review in this issue by Lindenbach and Rice, page 933), these findings cannot be directly extrapolated to viruses in infected individuals. Nonetheless, these *in vitro* data demonstrate the magnitude of the problem and allow an estimate of the genetic barrier that the virus has to overcome to acquire resistance to a given antiviral agent. Ultimately, it is likely that the combination of multiple drugs, possibly directed at viral as well as at host targets, will be required to contain the emergence of drug-resistant HCV variants and efficaciously treat chronic HCV infection.

This review provides an overview of recent progress in the identification of new targets and approaches. We will discuss, in this order, progress towards developing drugs targeting the viral enzymes or the viral genome, and then novel immunomodulatory molecules, with particular emphasis on the agents with demonstrated antiviral activity in the clinics.

Small-molecule inhibitors of viral enzymes

The development of direct antivirals that block essential viral enzymes represents a straightforward approach to developing new anti-HCV agents. Although all HCV enzymes are, in theory, equally appropriate for therapeutic intervention, the NS3-4A serine protease and the NS5B RNA polymerase have emerged as the most popular targets. A number of competitive inhibitors of the NS3 protease as well as nucleoside and non-nucleoside inhibitors of the NS5B polymerase are being developed. The efficacy shown by NS3 serine protease and the NS5B RNA-dependent RNA polymerase inhibitors in recent proof-of-concept clinical trials have validated the efforts spent in search of clinical candidates and triggered a renewed interest in this arena.

Inhibitors of the NS3-4A protease

The NS3-4A protease is a heterodimeric protease, comprising the amino-terminal domain of the NS3 protein and the small NS4A cofactor. Its activity is essential for the generation of components of the viral RNA replication complex (see the review in this issue by Lindenbach and Rice, page 933). Structurally, the HCV enzyme is a member of the chymotrypsin serine protease family, but its het-

¹Istituto di Ricerche di Biologia Molecolare P. Angeletti, Via Pontina km 30.600, 00040 Pomezia-Rome, Italy.

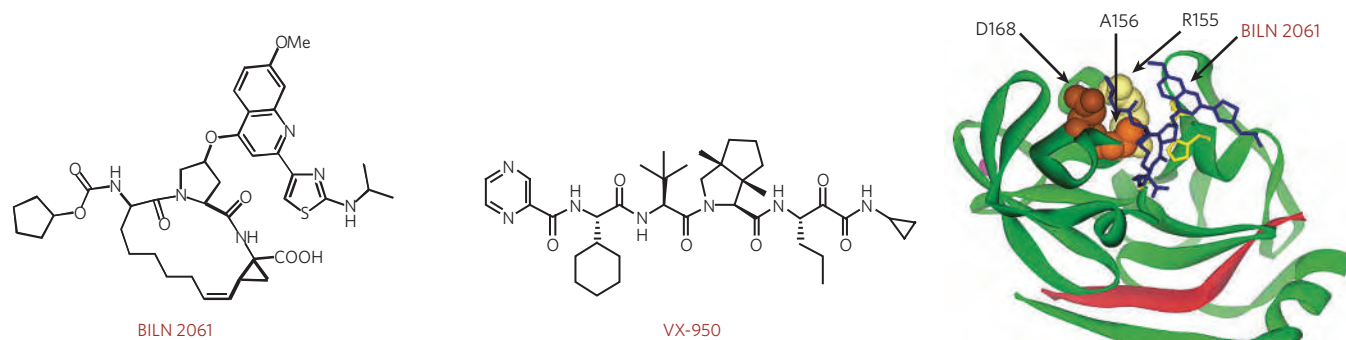


Figure 1 | The structures of inhibitors BILN 2061 and VX-950 and of the NS3 serine protease domain (green) complexed with the central domain of NS4A (red). The residues that constitute the enzyme catalytic triad (histidine 57, aspartate 81 and serine 139) are shown as yellow (stick representation). The structural zinc atom is indicated in purple. The

chemical structures of the protease inhibitors BILN 2061 and VX-950 are shown on the left. BILN 2061 is also modelled in the active site. Residues responsible for resistance to peptidomimetic protease inhibitors are shown in space-filling mode and colour-coded. See text for details.

erodimeric nature and the presence of a structural zinc atom differentiate the NS3-4A protease from the other members of the family (Fig. 1). The enzyme cleaves the viral polyprotein at four junctions with a temporal sequence that is presumably crucial for replication. In addition, the NS3-4A protease activity has been implicated in blocking the host cell's ability to mount an innate antiviral response (see the review in this issue by Gale and Foy, page 939). This observation has fostered the hope that inhibitors of this enzyme will block virus replication with a double hit, leading to an increased antiviral efficacy.

The substrate specificity of the NS3-4A protease is very different from that of related host enzymes⁴. Although this made the identification of inhibitors for the HCV enzyme conceivable, early biochemical and structural studies provided the first glimpse that developing such agents as drugs was no trivial task. The enzyme requires a long peptide substrate, with which it establishes multiple weak interactions distributed along an extended surface. The requirement for such long substrates instilled the fear that the enzyme could only be inhibited by molecules sufficiently large to mimic the natural substrate, and large molecules cannot be easily converted into drugs. The resolution of the three-dimensional structure of the enzyme provided another blow to the enthusiasm of drug developers. The substrate-binding cleft of NS3-4A protease seemed flat and featureless, lacking the cavities, holes and flaps that had been exploited as anchor points to design potent and selective inhibitors of other proteases^{5–7}. In spite of these difficulties, a few groups continued their search for inhibitors capitalizing on the observation that the enzyme is susceptible to marked inhibition by the N-terminal peptide products released from the substrates upon enzymatic cleavage^{8,9}.

BILN 2061 (Ciluprevir; Boehringer Ingelheim), a macrocyclic mimic of peptide product inhibitors (Fig. 1), was the first HCV protease inhibitor to enter clinical trials¹⁰. The development of BILN 2061 was built upon the finding that the carboxy-terminal carboxylic acid of the hexapeptide products provided an active-site affinity anchor and a moiety that warranted the desired selectivity for the HCV enzyme over the cellular proteases¹¹. Structural information guided optimization of the side chains and the conversion of the linear peptidic leads into macrocyclic inhibitors with enhanced potencies and improved biopharmaceutical properties^{12,13}, culminating in the selection of BILN 2061 as a clinical candidate. When administered to patients infected by genotype-1 HCV twice a day for only two days, BILN 2061 induced a rapid, dose-dependent decline of the viral load, exceeding a 2 log₁₀ reduction in all individuals receiving the higher doses of the drug^{10,14}. Treatment with BILN 2061 was somewhat less effective on genotype-2 and -3 viruses, although it resulted in a clinically meaningful reduction of the viral titre in several patients¹⁵. Although the effect of BILN 2061 was transient, the hope is that longer treatments with HCV protease inhibitors will achieve high rates of sustained viral

response and particularly so in patients that do not respond well to IFN therapy.

Unfortunately, the clinical development of BILN 2061 was halted because of the observation of cardiac toxicity in laboratory animals¹⁵. Toxicity may not be the only concern, however. As this agent progressed to clinical trials, development of resistance to BILN 2061 was observed in the HCV replicon system^{16,17}. A single mutation is sufficient to confer resistance to BILN 2061 *in vitro*, predicting that escape mutants might emerge with high frequency in HCV-infected patients. Substitution for either arginine 155, alanine 156 or aspartate 168 in the NS3 protease domain (Fig. 1) induced a high degree of resistance to BILN 2061 and to related analogues^{16–18}. Intriguingly, replacement of aspartate 168 with glutamine is naturally found in the NS3 protein of genotype 3 viral isolates, possibly explaining the reduced activity of BILN 2061 strain patients infected with this viral strain¹⁹.

VX-950 (Vertex/Mitsubishi) is a product-derived peptidomimetic inhibitor of the NS3-4A protease that is stabilized into the enzyme's

Box 1 | Laboratory tools for evaluating HCV antiviral resistance

Historically, the lack of readily available animal models for HCV infection and the inability to infect cultured cells have been major obstacles in establishing reliable antiviral assays. A major breakthrough for evaluating candidate antiviral agent in cell culture is represented by the so-called HCV replicons. HCV replicons are engineered subgenomic HCV RNAs capable of autonomous replication once introduced in cultivated cells.

The lack of a robust laboratory infection system makes preclinical studies of HCV antiviral resistance also quite challenging. However, it has been recently shown that the HCV replicon system, combined with biochemical assays and reverse genetics, can be exploited to study the emergence of resistance to anti-HCV agents *in vitro*.

This type of *in vitro* evolution and selection is made possible by two replicon features. On one hand, the low fidelity of RNA replication in tissue culture generates a genetic diversity in the replicon population that is likely to recapitulate the heterogeneity of the natural HCV quasispecies. On the other hand, the inclusion of a dominant selectable marker in the replicon sequence, such as neomycin phosphotransferase, allows the selection of replicon variants that become insensitive to the antiviral agent used during the selection. In brief, the ability of the replicon cells to survive in the presence of the antibiotic neomycin relies on replicon-driven expression of neomycin phosphotransferase. Thus, inhibitors of viral RNA replication abolish expression of resistance to neomycin in cells that harbour wild-type replicons. Under these conditions, cells containing replicons sensitive to the inhibitor are eliminated because of the antibiotic effect, while cells containing replicon variants with decreased sensitivity to the antiviral agent under study will survive and give rise to clones that can be isolated and characterized. Taking advantage of these features, it is possible to isolate cell clones containing resistant replicons by culturing a relatively small number of cells in the presence of neomycin and of a given inhibitor.

active site through the inclusion of an α -ketoamide²⁰ (Fig. 1). The α -ketoamide moiety is an enzyme active-site anchor capable of forming a reversible covalent bond with the catalytic serine. The incorporation of such a 'warhead' is a customary approach to making potent serine-protease inhibitors but is often disregarded because covalent protein binding is potentially responsible for an unfavourable safety profile. Not only did VX-950 efficiently inhibit HCV replication in cell culture, it was also sufficiently tolerated in laboratory animals to prompt its advancement in clinical trials. Interim results of phase Ia/Ib clinical trials indicate that VX-950 was well tolerated and demonstrated outstanding antiviral activity²¹. Treatment with VX-950 induced a fast decline of the viral load in patients infected by genotype-1 HCV, with a median reduction in HCV RNA ranging between 2 and 4.4 log₁₀ at the end of 2 weeks of therapy. Although this relatively short treatment was not sufficient to eradicate the virus, and viraemia returned to baseline after stopping therapy, these encouraging results have stimulated the planning of additional clinical studies.

VX-950 interacts covalently with the protease but this does not render it immune to the development of viral resistance¹⁷. There is only partial cross-resistance between BILN 2061 and VX-950, possibly reflecting the differences in the interaction of different inhibitor types with the enzyme. In fact, substitutions for aspartate 168 of NS3 confer resistance to BILN 2061 but not to VX-950. Conversely, replacement of alanine 156 yields different outcomes depending on the nature of the mutation. Thus, replacement of alanine 156 with serine confers selective resistance to VX-950 (ref. 17), whereas replacement of the same residue with threonine or valine confers significant cross-resistance to VX-950 and BILN 2061 as well as to structurally different protease inhibitors^{22,23}.

Several other peptidomimetic inhibitors of the NS3-4A protease are at various stages of development. However, since cross-resistance to diverse protease inhibitors can be achieved with the mutation of a single amino acid, the combination of multiple protease inhibitors could turn out to have limited value in the clinical practice.

NS5B polymerase inhibitors

The RNA-dependent RNA polymerase (RdRp) contained within the NS5B protein is the catalytic component of the HCV RNA replication machinery²⁴. This enzyme synthesizes RNA using an RNA template.

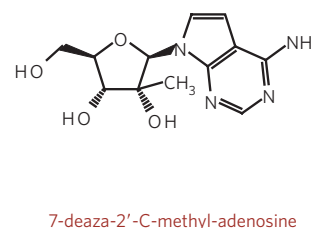
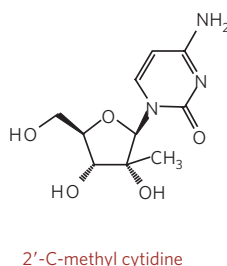
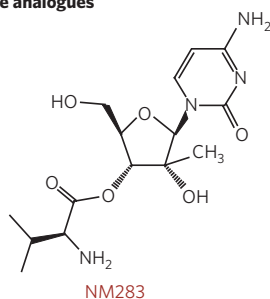
This biochemical activity is not present in mammalian cells, offering the opportunity to identify very selective inhibitors of the viral enzyme. In common with other polymerases, the structure of the NS5B enzyme is imaginatively assimilated to that of a right hand^{25–27} (Fig. 3). The palm domain contains the active site of the enzyme whereas the fingers and the thumb modulate the interaction with the RNA chain. In the conformation more frequently seen in the crystal, two loops extending from the fingers domain — the fingertips — are in contact with the thumb domain, leading to the formation of a 'closed' active site tunnel in which the RNA template, the nascent RNA strand and the nucleotide substrates are accommodated during the polymerization reactions. The NS5B polymerase can also adopt a more 'open' conformation, where the contact between the fingertips and the thumb is disrupted, resulting in a wider access to the catalytic site²⁸. As noted below, the resolution of the structure in complex with allosteric inhibitors suggests that some compounds work by freezing the enzyme in the 'open', inactive conformation.

Unlike the NS3-4A protease, the NS5B RdRp is a very tractable drug discovery target. The rational search of substrate analogues led to the identification of several nucleoside analogues, whereas high-throughput screening efforts have uncovered a variety of non-nucleoside inhibitors (NNIs) (Fig. 2). Nucleoside analogues need to be converted by the host-cell machinery to the corresponding nucleotides, which in turn inhibit synthesis of viral RNA as 'chain terminators', that is, substrate analogues that are incorporated by the viral polymerase in the nascent RNA molecule and induce premature termination of the RNA synthesis. Conversely, NNIs are almost invariably allosteric inhibitors believed to block the enzyme, preventing a conformational transition needed for initiation of RNA synthesis²⁹. Interestingly, several different binding sites for NNIs exist on the HCV polymerase. Furthermore, the different classes of nucleoside analogues and NNIs elicit diverse patterns of resistance. Thus, although these agents target the same viral enzyme, they offer the potential to be developed for combination therapy.

Nucleoside analogues

NM283 (Valopicitabine; Idenix/Novartis) is so far the only inhibitor of the NS5B polymerase with demonstrated antiviral activity in the clinics³⁰. NM283 is an oral prodrug of 2'-C-methyl-cytidine (Fig. 2). The

a Nucleoside analogues



b Non-nucleoside inhibitors

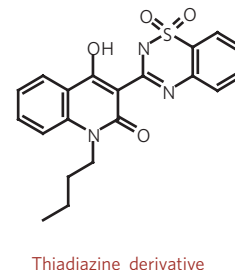
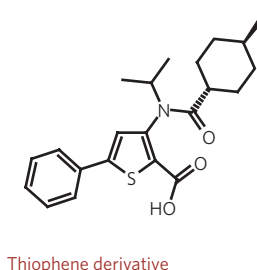
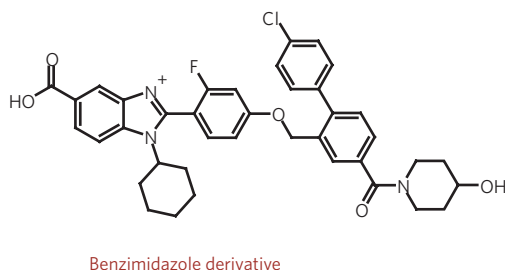


Figure 2 | Examples of inhibitors of the HCV RNA-dependent RNA polymerase. a, Structures of 2'-C-methyl-nucleosides. NM283 is a valine ester of NM107. **b,** Structures of representative examples of different classes

of non-nucleoside inhibitors of the NS5B polymerase, each presumably binding at a different allosteric site on the enzyme: benzimidazoles, benzothiadiazines and thiophenes.

latter compound was initially identified as an inhibitor of the HCV-related bovine viral diarrhoea virus (BVDV) and later shown to inhibit HCV RNA replication in the replicon assay^{31,32}. Interestingly, 2'-C-methyl purine nucleosides (Fig. 2) also inhibit replication of HCV, BVDV and other Flaviviruses, indicating that the addition of a methyl group at the 2' position of the ribose is sufficient to transform the nucleotide substrates into specific chain-terminators of the *Flaviviridae* RdRps^{33,34}. When administered to genotype-1 HCV patients for at least 2 weeks, NM283 induced a dose-dependent decline of the viral load to less than 10% of the initial levels in patients receiving an 800 mg daily dose³⁰. As expected, in all patients viraemia returned to pre-treatment levels after stopping therapy. Although NM283 was not as effective or rapid as BILN 2061 and VX-950, it demonstrated antiviral activity at tolerated doses. This encouraged the progression to longer term clinical trials in combination with pegylated IFN- α . Interim results after 24 weeks of therapy show that this combination caused a mean reduction in viral load of more than 10,000-fold, and HCV RNA was undetectable in a significant fraction of the patients³⁵. More extensive evaluation in combination trials is needed to assess whether it will achieve a sustained viral response rates superior to the current standard of care.

2'-C-methyl-nucleosides inhibit NS5B enzymes and replicons derived from different HCV genotypes and may have potential for treatment for all viral strains^{32,36}. But these agents are also readily susceptible to resistance development^{33,34}. Selection of replicons resistant to 2'-C-methyl-nucleosides has shown that HCV quickly learns to discriminate between these agents and the natural nucleotides. The virus acquires resistance to 2'-C-methyl-nucleosides by replacing serine 282 of NS5B with threonine. Ironically, from a chemical viewpoint, this mutation corresponds to the addition of a methyl group, the same modification used to convert the enzyme substrates into inhibitors. Replicons carrying the replacements of serine 282 with threonine show a decreased replication fitness³³. It remains to be defined whether this mutation will have similar debilitating effects on HCV replication also in a more physiological setting.

Non-nucleoside inhibitors

Among the NNIs of the NS5B polymerase, at least three different classes of compound are being evaluated in the clinic, but limited data have been disclosed on these drug candidates. The first NNIs of the NS5B polymerase to enter clinical trials were two oral agents JTK-109 and JTK-003 (Japan Tobacco). Very little has been so far disclosed on the clinical development of these compounds except that JTK-003 has been advanced to phase II. From the patent literature³⁷, JTK-109 and JTK-003 are assumed to be part of a heterogeneous series of 6,5-fused heterocyclic compounds based on a benzimidazole or indole core (Fig. 3). These compounds act as allosteric inhibitors and block polymerase activity before elongation, presumably impeding the conformational transition needed for the formation of a productive polymerase-RNA complex^{38,39}.

The very recent determination of the structure of the polymerase in complex with related analogues showed that these inhibitors bind on the surface of the thumb domain in a cavity that in the free enzyme is normally occupied by one of the fingertips (Fig. 3)⁴⁰. Inhibitor binding disrupts the fingertips-thumb interactions and forces the enzyme into an 'open', inactive conformation. Resistance to this class of inhibitor arises through a single mutation within the inhibitor binding site — replacement of proline 495 with alanine or leucine — which strongly reduces affinity for the inhibitors³⁹. Replicons carrying substitutions for proline 495 replicate inefficiently, but their fitness can be restored by mutations elsewhere in the NS5B coding region. In spite of the observation that proline 495 is absolutely conserved across all HCV isolates, compounds with this mechanism of action are significantly less active on enzymes and replicons derived from genotype-2 clinical isolates³², potentially limiting their clinical use.

A number of other polymerase NNIs progressed to clinical trials: these include R803 (Rigel) and HCV-371, HCV-086 and HCV-796

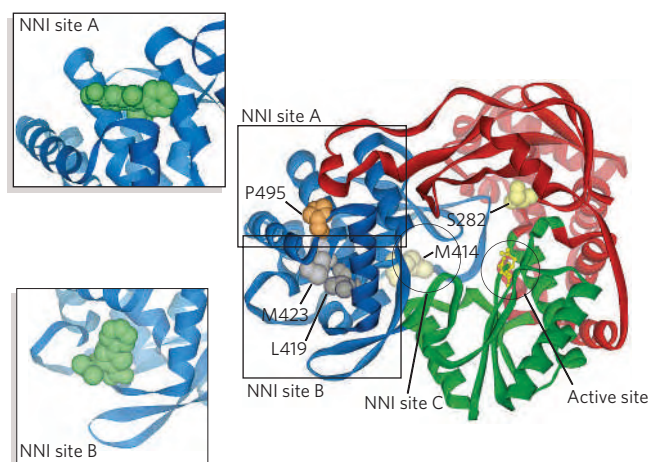


Figure 3 | The crystal structure of the NS5B RNA-dependent RNA polymerase. The thumb, palm and fingers domains are coloured in blue, green and red, respectively. The side chains of the two catalytic aspartates are in yellow in the active site (stick representation). The residues responsible for resistance to 2'-C-methyl nucleoside analogues (serine 282), benzimidazole (proline 495), thiophene (methionine 419 or threonine 423) or thiazidine inhibitors (methionine 414) are shown in space-filling style and indicated in the figure. The insets show the details of the polymerase in complex with a 5,6-fused heterocyclic inhibitor (NNI site A) or a thiophene inhibitor (NNI site B). For clarity, space-filling models of the inhibitors are shown in the insets in bright green. In the apo-protein structure, the thiophene binding pocket is occupied by the solvent while the benzimidazole binding site is occupied by the short α -helix at the extremity of one of the fingertip loops. NNI site C indicates a tentative location of the binding site for the inhibitors of the thiazidine class based on resistance and indirect biochemical data.

(ViroPharma/Wyeth). R803 was identified and optimized using the replicon assay⁴¹. Subsequently, biochemical and resistance studies demonstrated that R803 and related inhibitors targeted the NS5B polymerase. In fact, substitutions for tyrosine 452 or arginine 465 of the NS5B polymerase conferred resistance to a close analogue of R803. The unique resistance pattern of this class of compound suggests that they are mechanistically distinct from other known NNIs. R803 was selected for clinical development, but the clinical investigation was soon terminated because of poor efficacy of the compound, presumably because of inadequate pharmacokinetics⁴². Related compounds are being developed, including a prodrug of R803 and derivatives of R803 with improved pharmacokinetic properties.

No data have been disclosed yet on the preclinical characterization of HCV-371, HCV-086 and HCV-796, but they are allegedly allosteric inhibitors of the NS5B polymerase. Development of HCV-371 and HCV-086 was also halted because of insufficient antiviral activity. Nonetheless, the data obtained with HCV-086 expedited clinical studies of the related compound HCV-796, a more potent analogue of HCV-086 with demonstrated antiviral activity in an animal model of hepatitis C infection⁴³.

Aside from the compounds that have made it to the clinic, several other NS5B NNIs have been reported to inhibit HCV replication in tissue culture and are being considered for development²⁹. Among these, a class of thiophene derivatives has been characterized extensively^{44,45}. These compounds are reversible allosteric inhibitors of the NS5B polymerase and bind to the enzyme at the base the thumb domain, in a long cleft that is close to, but distinct from, the site occupied by benzimidazole-based inhibitors²⁸ (Fig. 3). Intriguingly, binding of thiophene inhibitors also induces a conformational change of the enzyme toward a more 'open' form, suggesting that the two classes of compound inhibit RNA synthesis with a similar mechanism. *In vitro* studies have shown that resistance to thiophene inhibitors arises through substitution for either of two residues — methionine 419 or threonine 423 — located in the hydrophobic region of the inhibitors binding site⁴. Interestingly, although the thiophene-binding region is

well conserved among HCV genotypes, compounds of this class are significantly less potent on polymerases of non-1 genotypes³². The crystallographic evidence that the same binding pocket is also targeted by structurally different inhibitors based on a phenylalanine or dihydropyranone scaffold provides important information for further optimization of inhibitors binding at this site^{46,47}.

Another class of allosteric inhibitor of the HCV polymerase is based on a benzothiadiazine scaffold⁴⁸ (Fig. 2). Similarly to other NNIs of the NS5B polymerase, compounds from this class inhibit RNA synthesis acting before the formation of an elongation complex^{49,50}. The available biochemical data, however, suggest that benzothiadiazines bind the enzyme at a different site and possibly act through different mechanisms from the allosteric inhibitors described above⁵⁰. This hypothesis was strengthened by the unique pattern of resistance mutations obtained in the replicon system with these compounds^{50,51}. Indeed, several different single NS5B mutations, mapping to different regions of the polymerase, have been reported to induce resistance to benzothiadiazines in the replicon system. Of these, only the mutants of methionine 414 demonstrated a clear resistance to inhibition in biochemical assays using the purified enzyme. The observed resistance could be ascribed to reduced affinity for the inhibitor, thus implicating methionine 414 as part of the inhibitor-binding site. On the basis of the location of this residue in the inner surface of the thumb domain, close to the active site, it is tempting to speculate that the binding site of benzothiadiazines is distinct from those of other allosteric inhibitors (Fig. 3). Sequence analysis of natural HCV isolates revealed that the NS5B residues involved in resistance to benzothiadiazines are not entirely conserved among HCV genotypes, implying that some naturally occurring HCV isolates might be resistant to this class of inhibitor. In line with this consideration, representative benzothiadiazines have been found to inhibit only a limited subset of enzymes and replicons derived from clinical isolates of different genotypes^{32,36}.

Nucleic-acid-based antiviral agents

The concept of using synthetic nucleic acids as drugs has received increasing attention over the past few years. In particular, antisense oligonucleotides, ribozymes and, more recently, siRNAs are being explored as therapeutic agents in a number of areas. One of the major issues that will determine the success of nucleic-acid-based drugs is efficient delivery of the synthetic polymers to the appropriate cells *in vivo*. Because of their size and chemical nature, nucleic-acid polymers are not orally bioavailable and can be administered only parenterally. Perhaps reflecting anatomical and physiological features, systemic administration seems to work best for the liver and less efficiently for other organs⁵². In this respect, HCV could be viewed as the ideal target for successful development of effective nucleic-acid-based drugs.

The HCV internal ribosome entry site (IRES) found at the 5' end of the viral genome has been considered the most attractive target for the development of RNA-based drugs, both because of the wealth of data available on IRES structure and function and because of the conservation among HCV genotypes. Ribozymes and antisense oligonucleotides designed to target the HCV IRES reduced the HCV RNA translation and replication in cell culture. The ribozyme RPI.13919 (Heptazyme; RPI) and the antisense oligonucleotide ISIS-14803 (ISIS Pharmaceuticals) have both progressed to early-phase clinical studies in HCV-infected patients⁵³, but their development was halted because of adverse effects or because of the limited efficacy.

RNA interference approach to HCV antivirals

RNA interference (RNAi) is the latest addition to the list of nucleic-acid based approaches being explored for HCV therapy^{54–58} (Box 2). In tissue culture, siRNA, as well as vector-encoded short hairpin RNA (shRNA) directed against the viral genome, effectively blocks the replication of HCV replicons. Not all siRNAs showed the same efficiency, but all regions of the HCV genome were susceptible to RNAi, including conserved sequences. The extent and the duration of silencing var-

ied greatly between the different studies, but the most effective siRNA seemed capable of completely eradicating HCV from more than 98% of the replicon-bearing cells⁵⁸. The positive outcome of these *in vitro* studies stimulated scientists to take up the challenge of identifying appropriate means for efficient *in vivo* delivery of siRNA. Given the notorious lack of readily accessible laboratory models for HCV infection, initial animal studies simply addressed the ability to knockdown messenger RNA expression in mouse liver. High-pressure injection in the tail vein of siRNA or shRNA targeting the surface receptor FAS, caspase 8 or a reporter gene containing HCV sequences silenced the cognate mRNAs^{59–61}. Despite studies proving that naked, unmodified siRNA or shRNA can effectively induce RNAi *in vivo*, it is doubtful that the hydrodynamic injection method employed has the potential for clinical application. Possibly more promising is the use of chemically modified siRNA. Several avenues are being explored and the most promising seem capable of enhancing siRNA stability and cell penetration without reducing the silencing efficiency. Interestingly, chemically modified siRNA seem able to inhibit HCV replication not only in tissue culture but also in a mouse model of HCV infection⁶².

Resistance development is a potential obstacle also for RNAi-based therapy. HCV can develop resistance to prolonged treatment with siRNA through the accumulation of nucleotide point mutations within the siRNA target sequence⁶³. As expected, replicons resistant to a given siRNA remain susceptible to siRNAs targeting different HCV RNA sequences, and the emergences of resistant replicons is diminished by the combination of two or more siRNAs. Thus, the use of two or more siRNAs targeting different sequences of the viral genome may provide a way to control the development of resistance. Indeed, BLT-HCV (Benitec), the first clinical candidate to treat HCV infection through RNAi, consists of three components targeting different HCV sequences, underlining the importance of a multi-targeting approach to prevent resistance development⁶⁴.

Novel immunomodulatory agents

The experience with IFN-based therapies has demonstrated that HCV infection can be eradicated by agents that stimulate the host innate and adaptive immunity. Fuelled by this observation, synthetic agonists of Toll-like receptors (TLRs) 7 and 9 have progressed through early-phase clinical trials and have now begun to show their potential in controlling HCV infection.

Toll-like receptors are molecular sentinels that sense the presence of invading microorganisms through the recognition of molecular patterns characteristic of pathogens such as bacteria, viruses and parasites⁶⁵. They are expressed by immune cells, which include macrophages, monocytes, dendritic cells and B cells⁶⁶. Signalling by stimulated TLRs initiates acute inflammatory responses by induction of antimicrobial genes and pro-inflammatory cytokines and chemokines.

So far, ten different TLRs have been identified in humans, each recognizing molecular patterns associated with a specific class of microbial agents⁶⁶. TLRs 3, 7, 8 and 9 are intracellular receptors specialized in the recognition of viral nucleic acids: they detect, respectively, double-stranded RNA (TLR-3), single-stranded RNA (TLR-7 and -8), and unmethylated deoxycytosine-deoxyguanosine (CpG) DNA sequences (TLR-9) in late endosomes and lysosomes⁶⁷. Recognition of ligands by cognate TLRs expressed on cells that are effectors of the innate immune response leads to the rapid activation of inflammation and microbicidal pathways⁶⁶. In addition, owing to their ability to induce the activation of antigen-presenting dendritic cells, a subset of TLRs can initiate T-cell priming and contribute to the establishment of an adaptive immune response⁶⁸. Signalling through TLRs also induces the production of T-helper (T_H) cells type 1, promoting cytokines and chemokines, steering the course towards the T_H1 phenotype, as needed to combat chronic viral infections and favour viral clearance. Stimulation of the appropriate TLRs may help restore the innate and adaptive immune functions that are dysfunctional in hepatitis-C-infected hosts⁶⁹. As discussed below, the immune response observed

Box 2 | Therapeutic RNA interference

RNA interference (RNAi) is a natural process used by eukaryotic cells to recognize and destroy abnormal or exogenous RNA⁷⁶. RNAi is triggered by the presence of RNA molecules containing double-stranded regions (dsRNA) which are recognized as exogenous and chopped into 21–23-nucleotide long duplexes, designated short interfering RNA (siRNA). siRNA associate with a number of cellular proteins to form an RNA-induced silencing complex (RISC) which recognizes RNA complementary to either of the siRNA strands and catalyses its cleavage.

The existence of RNAi in mammalian cells was initially masked by the fact that long dsRNA activate the interferon pathway resulting in a non-specific shutdown of translation. Only in 2001 did scientists discover that it was possible to avoid the interferon response and activate gene-specific RNAi by directly feeding siRNA to mammalian cells⁷⁷, opening the way to the therapeutic application of RNAi. The initial hype over the therapeutic use of RNAi was based on its potential to overcome some of the shortcomings of the previous antisense technology, the most important advantage being the catalytic nature of RNAi. Unfortunately, soon it became evident that RNAi was not exempt from the quandaries observed with antisense, and in particular from the problems connected with *in vivo* delivery and stability. Chemical modification of the siRNA is emerging as a valid approach to improve stability and permeability characteristics⁷⁸. Alterations of the phosphodiester-ribose backbone, such as the addition of 2'-O-methyl groups, have been shown to increase the siRNA resistance to nucleases and prolong the silencing effect. Likewise, various tricks have been used to enhance cell penetration both in cell culture and *in vivo*, one of the most popular being the conjugation with lipophilic molecules such as cholesterol. A different solution to the delivery problems is the use of vectors expressing siRNAs in the form of short hairpin RNA (shRNA) that are converted into siRNA by the cellular processing machinery.

during the course of acute viral infections can be mimicked by the use of synthetic agonists of TLRs.

Antiviral activity of TLR-9 and TLR-7 agonists

TLR-9 and TLR-7, in humans, are both expressed in B cells and in plasmacytoid dendritic cell (PDCs). In addition, TLR-7 is expressed in myeloid dendritic cells⁶⁶. PDCs are the key effectors in the innate immune response to invading viruses because of their extraordinary capacity to produce very high levels of type 1 (α and β) IFNs in response to TLR-9 and TLR-7 stimulation by the viral nucleic acids⁷⁰. As a consequence of the secretion of type 1 IFNs, a number of secondary effects are induced that link the innate to the adaptive immune response, such as stimulation of natural killer (NK) cells as well as maturation of PDCs to potent antigen-presenting cells.

Short synthetic oligonucleotides containing one or more unmethylated CpG motifs flanked by specific sequences are potent agonists of TLR-9 (ref. 71). Stimulation of PDCs with such oligonucleotides results in the production of tumour necrosis factor- α (TNF- α), interleukin (IL)-12 and high levels of IFN- α . Additionally, TLR-9 ligands are potent stimulators of B-cell proliferation and antibody secretion.

A CpG-containing oligonucleotide, CPG-10101 (Actilon; Coley Pharmaceutical Group), was evaluated in HCV-infected patients, and yielded promising results⁷². The drug was administered subcutaneously twice weekly over a period of four weeks to individuals with chronic HCV infection who had failed previous IFN-based therapy. One-third of the patients demonstrated early viral level reduction of at least an order of magnitude during the treatment. The viral level reduction observed was consistent with the elevation of IFN- α and other markers associated with an antiviral immune response, thus validating the rationale for exploiting TLR-9 agonists for the treatment of chronic HCV infection.

TLR-7 recognizes several synthetic compounds that are structurally related to nucleic acids. These include imidazoquinolines, loxoribine (7-allyl-7,8-dihydro-8-oxoguanosine), and broprimine (2-amin-5-bromo-6-phenyl-4(3)-pyrimidinone)⁶⁶. The antiviral properties of synthetic TLR-7 ligands have been long known: these molecules induce a potent, broad-spectrum antiviral response owing to their ability to induce the release of inflammatory cytokines, especially IFN- α .

ANA245 (7-thia-8-oxoguanosine or Isatoribine; Anadys Pharmaceuticals) is a guanine nucleoside analogue whose immunostimulatory activity depends on its agonistic activity on TLR-7 (ref. 73). To provide the evidence that an agonist of TLR-7 could show anti-HCV activity, ANA245 was administered subcutaneously to a small group of HCV patients infected with different HCV genotypes⁷⁴.

Patients administered the highest daily dose of ANA245 for a week showed a statistically significant reduction in viral load, with some of them achieving reduction of more than 90% at the end of the treatment. Although these results are preliminary, they provide proof that a compound interacting with TLR-7 can reduce viral load in HCV-infected patients. In addition, the patient responses did not appear to be dependent on the viral genotype. Following the encouraging results in clinical study, ANA975, an oral prodrug of ANA245, is currently being developed⁷⁵. If successful, ANA975 could combine the broad-spectrum efficacy of an immune-based therapy with the convenience of administration of an oral drug.

Outlook and future challenges

Several novel drugs have entered or will soon enter clinical evaluation to establish their clinical usefulness for HCV patients. Aside from the safety and efficacy requirements common to all new drugs, the success of HCV-targeted agents will be heavily influenced by their ability to inhibit all viral variants and prevent the emergence of escape mutants. Although agents targeting host rather than viral factors are less likely to fail owing to this problem, no drug can be considered totally exempt from the risk of resistance development. As is the case for HIV, combinations of several antiviral agents attacking different viral and possibly host targets will almost certainly be required to control infection and prevent the emergence of drug-resistant viral variants.

Although the first wave of experimental HCV-targeted drugs is yielding promising results in early clinical trials, a wider repertoire of

Table 1 | A sample of the drug pipeline for hepatitis C

a HCV-targeted drugs				
Compound name(s)	Company	Clinical phase	Target	Mechanism of action
BILN 2061 (Ciluprevir)	Boehringer-Ingelheim	Phase II*	NS3-4A protease	Product-derived serine protease inhibitor
VX-950	Vertex/Mitsubishi	Phase Ib	NS3-4A protease	Serine protease reversible covalent inhibitor
NM283 (Valopicitabine)	Idenix/Novartis	Phase II	NS5B polymerase	Nucleoside analogue (chain terminator)
JTK-103	Japan Tobacco	Phase II	NS5B polymerase	Non-nucleoside allosteric inhibitor
HCV-796	ViroPharma/Wyeth	Phase Ia	NS5B polymerase	Non-nucleoside allosteric inhibitor
*Development halted due to cardiotoxicity in monkeys				
b Host targets/immunomodulators				
Actilon (CpG-10101)	Coley Pharmaceutical Group	Phase Ib	Toll-like receptor-9	Immunomodulator
ANA245 (Isatoribine)	Anadys Pharmaceuticals	Phase Ib	Toll-like receptor-7	Immunomodulator
ANA975	Anadys Pharmaceuticals	Phase Ia	Toll-like receptor-7	Immunomodulator (prodrug of ANA245)

novel antiviral agents is needed. For this, drug discoverers must consider and pursue all other known viral targets for which there are not yet clinically useful inhibitors. These include the p7 ion channel, the NS2-3 cysteine protease and the NS3 helicase. Moreover, the advancement in the field of therapeutic RNAi makes it conceivable to target a variety of host factors that are essential for viral replication or persistence and for which development of small-molecule inhibitors may turn out to be prohibitively difficult.

The past clinical experience has demonstrated that, when IFN-based therapies are successful, the hepatitis C virus can be permanently cleared from the host tissues. Thus, the goal for future therapeutic regimens will be that of achieving a complete and sustained viral clearance in as many patients as possible.

What combination will turn out to be successful for HCV remains an open question. After initial proof-of-concept demonstration of antiviral activity, most new drugs are being tested in combination with IFN- α or IFN- α and ribavirin. In the future, as more agents become available, it will become possible to conceive combination schemes based entirely on novel drugs. Considering the limitations imposed by the severity of the side effects associated with the agents currently approved for hepatitis C, a shift away from IFN-based treatments would constitute a major breakthrough. A crucial question that future clinical studies need to address is whether combination therapy with solely HCV-targeted drugs will be sufficient to cure patients or whether the stimulation of the host immune system by immunomodulators or therapeutic vaccines will be necessary to completely eliminate the virus. ■

- Wasley, A. & Alter, M. J. Epidemiology of hepatitis C: geographic differences and temporal trends. *Semin. Liver Dis.* **20**, 1–16 (2000).
- Brown, R. S. Jr & Gaglio, P. J. Scope of worldwide hepatitis C problem. *Liver Transpl.* **9**, S10–S13 (2003).
- Tomei, L., Altamura, S., Paonessa, G., De Francesco, R. & Migliaccio, G. HCV antiviral resistance: the impact of *in vitro* studies on the development of antiviral agents targeting the viral NS5B polymerase. *Antiviral Chem. Chemother.* **16**, 225–245 (2005).
- Urbani, A. *et al.* Substrate specificity of the hepatitis C virus serine protease NS3. *J. Biol. Chem.* **272**, 9204–9209 (1997).
- Yan, Y. *et al.* Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form. *Protein Sci.* **7**, 837–847 (1998).
- Kim, J. L. *et al.* Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **87**, 343–355 (1996).
- Love, R. A. *et al.* The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* **87**, 331–342 (1996).
- Llinas-Brunet, M. *et al.* Peptide-based inhibitors of the hepatitis C virus serine protease. *Bioorg. Med. Chem. Lett.* **8**, 1713–1718 (1998).
- Steinkuhler, C. *et al.* Product inhibition of the hepatitis C virus NS3 protease. *Biochemistry* **37**, 8899–8905 (1998).
- Lamarre, D. *et al.* An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **426**, 186–189 (2003).
- Llinas-Brunet, M. *et al.* Highly potent and selective peptide-based inhibitors of the hepatitis C virus serine protease: towards smaller inhibitors. *Bioorg. Med. Chem. Lett.* **10**, 2267–2270 (2000).
- Goudreau, N. *et al.* NMR structural characterization of peptide inhibitors bound to the Hepatitis C virus NS3 protease: design of a new P2 substituent. *J. Med. Chem.* **47**, 123–132 (2004).
- Tsantrizos, Y. S. *et al.* Macrocyclic inhibitors of the NS3 protease as potential therapeutic agents of hepatitis C virus infection. *Angew. Chem. Int. Ed. Engl.* **42**, 1356–1360 (2003).
- Hinrichsen, H. *et al.* Short-term antiviral efficacy of BILN 2061, a hepatitis C virus serine protease inhibitor, in hepatitis C genotype 1 patients. *Gastroenterology* **127**, 1347–1355 (2004).
- Reiser, M. *et al.* Antiviral efficacy of NS3-serine protease inhibitor BILN-2061 in patients with chronic genotype 2 and 3 hepatitis C. *Hepatology* **41**, 832–835 (2005).
- Lu, L. *et al.* Mutations conferring resistance to a potent hepatitis C virus serine protease inhibitor *in vitro*. *Antimicrob. Agents Chemother.* **48**, 2260–2266 (2004).
- Lin, C. *et al.* *In vitro* resistance studies of hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061: structural analysis indicates different resistance mechanisms. *J. Biol. Chem.* **279**, 17508–17514 (2004).
- Trozzi, C. *et al.* *In vitro* selection and characterization of hepatitis C virus serine protease variants resistant to an active-site peptide inhibitor. *J. Virol.* **77**, 3669–3679 (2003).
- Thibeault, D. *et al.* Sensitivity of NS3 serine proteases from hepatitis C virus genotypes 2 and 3 to the inhibitor BILN 2061. *J. Virol.* **78**, 7352–7359 (2004).
- Chen, S.-H. *et al.* P1 and P1' optimization of [3,4]-bicycloproline P2 incorporated tetrapeptidyl α -ketoamide based HCV protease inhibitors. *Lett. Drug Des. Disc.* **2**, 118–123 (2005).
- Vertex Pharmaceuticals reports that oral hepatitis C protease inhibitor VX-950 dramatically reduces viral levels in phase Ib clinical study. <http://www.vrtx.com/Pressreleases2005/pr051705.html> (2005).
- Lemon, S. M., Yi, M. & Li, K. Strong reasons make strong actions. The antiviral efficacy of NS3/4A protease inhibitors. *Hepatology* **41**, 671–673 (2005).
- Lin, C. *et al.* *In vitro* resistance mutations against VX-950 and BILN 2061, two protease inhibitor clinical candidates: single-resistance, cross-resistance and fitness. *Hepatology* **40** (Suppl. 1), 404A (2004).
- Behrens, S. E., Tomei, L. & De Francesco, R. Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J.* **15**, 12–22 (1996).
- Ago, H. *et al.* Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Struct. Fold. Des.* **7**, 1417–1426 (1999).
- Bressanelli, S. *et al.* Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl Acad. Sci. USA* **96**, 13034–13049 (1999).
- Lesburg, C. A. *et al.* Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nature Struct. Biol.* **6**, 937–943 (1999).
- Biswal, B. K. *et al.* Crystal structures of the RNA dependent RNA polymerase genotype 2a of hepatitis C virus reveal two conformations and suggest mechanisms of inhibition by non-nucleoside inhibitors. *J. Biol. Chem.* **280**, 18202–18210 (2005).
- Beaulieu, P. L. & Tsantrizos, Y. S. Inhibitors of the HCV NS5B polymerase: new hope for the treatment of hepatitis C infections. *Curr. Opin. Investig. Drugs* **5**, 838–850 (2004).
- Afdhal, N. *et al.* Final phase I/II trial results for NM283, a new polymerase inhibitor for hepatitis C: antiviral efficacy and tolerance in patients with HCV-1 infection, including previous interferon failures. http://www.idenix.com/products/datapres_nm283/AfdhalAASLD04_10-04.pdf (2004).
- Standing, D. N. NM283 has potent antiviral activity against chronic hepatitis C virus, genotype 1, in the chimpanzee. http://www.idenix.com/products/datapres_nm283/StandingEASL2004_7-03.pdf (2004).
- Carroll, S. *et al.* Susceptibility of different genotypes of hepatitis C virus to inhibition by nucleoside and nonnucleoside inhibitors. *Antiviral Res.* **62**, A83 (2004).
- Migliaccio, G. *et al.* Characterization of resistance to non-obligate chain-terminating ribonucleoside analogs that inhibit hepatitis C virus replication *in vitro*. *J. Biol. Chem.* **278**, 49164–49170 (2003).
- Olsen, D. B. *et al.* A 7-deaza-adenosine analog is a potent and selective inhibitor of hepatitis C virus replication with excellent pharmacokinetic properties. *Antimicrob. Agents Chemother.* **48**, 3944–3953 (2004).
- Afdal, N. *et al.* Enhanced antiviral efficacy for Valopicitabine (NM283) plus peg-interferon in hepatitis C patients with HCV genotype-1 infection: Results of a phase IIa multicenter trial. http://www.idenix.com/products/datapres_nm283/AfdhalEASL2005_4-05.pdf (2005).
- Ludmerer, S. W. *et al.* Replication fitness and NS5B drug sensitivity of diverse hepatitis C virus isolates characterized by using a transient replication assay. *Antimicrob. Agents Chemother.* **49**, 2059–2069 (2005).
- Hashimoto, H., Mizutani, K. & Yoshida, A. in WO 00147883 (Japan Tobacco Inc., Published International Patent Application, 2001).
- LaPlante, S. R. *et al.* Binding mode determination of benzimidazole inhibitors of the hepatitis C virus RNA polymerase by a structure and dynamics strategy. *Angew. Chem. Int. Ed. Engl.* **43**, 4306–4311 (2004).
- Tomei, L. *et al.* Mechanism of action and antiviral activity of benzimidazole-based allosteric inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *J. Virol.* **77**, 13225–13231 (2003).
- Di Marco, S. *et al.* Interdomain communication in hepatitis C virus polymerase abolished by small-molecule inhibitors bound to a novel allosteric site. *J. Biol. Chem.* published online 13 June 2005.
- Lu, H. in WO 2005/000308 (Rigel Pharmaceuticals, USA. Published International Patent Application, 2005).
- Poor bioavailability results in insignificant clinical effects for Rigel R803 in phase I/II HCV trial. http://www.rigel.com/rigel/pr_1101094254 (2004).
- ViroPharma announces data from HCV-086 proof of concept study. <http://phx.corporate-ir.net/phoenix.zhtml?c=92320&p=irol-researchNewsArticle&ID=684145&highlight> (2005).
- Chan, L. *et al.* Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 2: tertiary amides. *Bioorg. Med. Chem. Lett.* **14**, 797–800 (2004).
- Chan, L. *et al.* Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 1: Sulfonamides. *Bioorg. Med. Chem. Lett.* **14**, 793–796 (2004).
- Wang, M. *et al.* Non-nucleoside analogue inhibitors bind to an allosteric site on HCV NS5B polymerase. Crystal structures and mechanism of inhibition. *J. Biol. Chem.* **278**, 9489–9495 (2003).
- Love, R. A. *et al.* Crystallographic identification of a noncompetitive inhibitor binding site on the hepatitis C virus NS5B RNA polymerase enzyme. *J. Virol.* **77**, 7575–7581 (2003).
- Dhanak, D. *et al.* Identification and biological characterization of heterocyclic inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *J. Biol. Chem.* **277**, 38322–38327 (2002).
- Gu, B. *et al.* Arresting initiation of hepatitis C virus RNA synthesis using heterocyclic derivatives. *J. Biol. Chem.* **278**, 16602–16607 (2003).
- Tomei, L. *et al.* Characterization of the inhibition of hepatitis C virus RNA replication by nonnucleosides. *J. Virol.* **78**, 938–946 (2004).
- Nguyen, T. T. *et al.* Resistance profile of a hepatitis C virus RNA-dependent RNA polymerase benzothiadiazine inhibitor. *Antimicrob. Agents Chemother.* **47**, 3525–3530 (2003).
- Braasch, D. A. *et al.* Biodistribution of phosphodiester and phosphorothioate siRNA. *Bioorg. Med. Chem. Lett.* **14**, 1139–1143 (2004).
- Foster, G. R. Past, present, and future hepatitis C treatments. *Semin. Liver Dis.* **24** (Suppl. 2), 97–104 (2004).
- Kronke, J. *et al.* Alternative approaches for efficient inhibition of hepatitis C virus RNA replication by small interfering RNAs. *J. Virol.* **78**, 3436–3446 (2004).
- Kapadia, S. B., Brideau-Andersen, A. & Chisari, F. V. Interference of hepatitis C virus RNA replication by short interfering RNAs. *Proc. Natl Acad. Sci. USA* **100**, 2014–2018 (2003).
- Yokota, T. *et al.* Inhibition of intracellular hepatitis C virus replication by synthetic and vector-derived small interfering RNAs. *EMBO Rep.* **4**, 602–608 (2003).
- Wilson, J. A. *et al.* RNA interference blocks gene expression and RNA synthesis from hepatitis C replicons propagated in human liver cells. *Proc. Natl Acad. Sci. USA* **100**, 2783–2788 (2003).

58. Randall, G., Grakoui, A. & Rice, C. M. Clearance of replicating hepatitis C virus replicon RNAs in cell culture by small interfering RNAs. *Proc. Natl Acad. Sci. USA* **100**, 235–240 (2003).
59. Song, E. *et al.* RNA interference targeting Fas protects mice from fulminant hepatitis. *Nature Med.* **9**, 347–351 (2003).
60. Zender, L. *et al.* Caspase 8 small interfering RNA prevents acute liver failure in mice. *Proc. Natl Acad. Sci. USA* **100**, 7797–7802 (2003).
61. McCaffrey, A. P. *et al.* RNA interference in adult mice. *Nature* **418**, 38–39 (2002).
62. Han, J. *et al.* Inhibition of HCV replication *in vivo* by nuclease-resistant siRNAs that are targeted to the liver. Presented at 11th International Symp. Hepatitis C Virus and Related Viruses (Heidelberg, Germany 2004).
63. Wilson, J. A. & Richardson, C. D. Hepatitis C virus replicons escape RNA interference induced by a short interfering RNA directed against the NS5b coding region. *J. Virol.* **79**, 7050–8 (2005).
64. Benitec announces clinical candidate for treatment of hepatitis C. <http://www.benitec.com/PRDownloads/Hepatitis%20C%20Clinical%20Candidate%20050905%20.pdf> (2005).
65. O'Neill, L. A. TLRs: Professor Mechnikov, sit on your hat. *Trends Immunol.* **25**, 687–693 (2004).
66. Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annu. Rev. Immunol.* **21**, 335–376 (2003).
67. Boehme, K. W. & Compton, T. Innate sensing of viruses by toll-like receptors. *J. Virol.* **78**, 7867–7873 (2004).
68. Iwasaki, A. & Medzhitov, R. Toll-like receptor control of the adaptive immune responses. *Nature Immunol.* **5**, 987–995 (2004).
69. Hahn, Y. S. Subversion of immune responses by hepatitis C virus: immunomodulatory strategies beyond evasion? *Curr. Opin. Immunol.* **15**, 443–449 (2003).
70. McKenna, K., Beignon, A. S. & Bhardwaj, N. Plasmacytoid dendritic cells: linking innate and adaptive immunity. *J. Virol.* **79**, 17–27 (2005).
71. Schetter, C. & Vollmer, J. Toll-like receptors involved in the response to microbial pathogens: development of agonists for toll-like receptor 9. *Curr. Opin. Drug Discov. Dev.* **7**, 204–210 (2004).
72. Coley reports results from phase I studies of Actilon™ for hepatitis C. http://www.coleypharma.com/coley/pr_1105025921 (2005).
73. Lee, J. *et al.* Molecular basis for the immunostimulatory activity of guanine nucleoside analogs: activation of Toll-like receptor 7. *Proc. Natl Acad. Sci. USA* **100**, 6646–6651 (2003).
74. Horsmans, Y. *et al.* Isatoribine, a Toll-like receptor 7 agonist, significantly reduced plasma viral load in a clinical proof-of-concept study in patients with chronic hepatitis C virus. *Hepatology* **40** (Suppl. 1), 282A (2004).
75. Anadys Pharmaceuticals announces selection of ANA975 as a development candidate for front-line treatment of chronic hepatitis C. <http://phx.corporate-ir.net/phoenix.zhtml?c=148908&p=irol-newsArticle&ID=575761&highlight> (2004).
76. Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
77. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
78. Manoharan, M. RNA interference and chemically modified small interfering RNAs. *Curr. Opin. Chem. Biol.* **8**, 570–579 (2004).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for material should be addressed to R.D.F. (raffaele_defrancesco@merck.com).

Prospects for a vaccine against the hepatitis C virus

Michael Houghton¹ & Sergio Abrignani²

The recent discovery of natural immunity to the hepatitis C virus and vaccine efficacy in the chimpanzee challenge model has allowed optimism about the development of at least a partly effective vaccine against this heterogeneous pathogen that is responsible for much of the chronic liver disease around the world. The immune systems of some infected individuals can spontaneously clear the virus, whereas other people need treatment with antivirals that work partly by stimulating humoral and cellular immune responses. Therefore, therapeutic vaccine strategies are also being pursued to improve treatment outcome.

A decade ago, an effective vaccination against the hepatitis C virus (HCV) was considered only a remote possibility. Three factors contributed to this: the high propensity of HCV to promote chronic persistent infections¹; evidence that convalescent humans and chimpanzees could be readily reinfected following re-exposure²; and the considerable genetic heterogeneity of this positive-stranded RNA virus³.

The situation today is more positive for two reasons. First, we now know that spontaneous eradication of the virus occurs in up to 50% of acute infections⁴ and that this viral clearance is associated with specific immune responses to the virus. Recapitulation of such immune responses by appropriate vaccination is therefore a realistic option. Second, clear evidence for at least some natural immunity has emerged recently in both humans⁵ and chimpanzees^{6–8}. (Chimpanzees are the only animal model available and develop only mild clinical sequelae.) Convalescent humans and chimpanzees are protected against re-exposure to the virus in the majority of cases, even against very divergent viral strains. Importantly, protection is usually at the level of prevention of progression to chronic, persistent infection following re-exposure rather than prevention of acute reinfection but this could translate to effective prophylaxis because, in humans, it is the chronic, persistent nature of HCV infection that is mainly associated with viral pathogenicity^{1,4}. Although some re-exposed individuals develop chronic infection⁹, most do not^{5–8}. This suggests that the generation of at least a partly effective vaccine against HCV is feasible. Indeed, emerging vaccine efficacy data from the chimpanzee challenge model indicate that it is possible to impede the progression to chronic infection in vaccinees. Until very recently^{10–12}, it was not possible to grow HCV efficiently in cell culture, and so the use of inactivated or live, attenuated viral vaccines has not yet been evaluated. Vaccine approaches have therefore included the use of adjuvanted recombinant polypeptide subunits of the virus in attempts to prime viral-neutralizing antibodies to the envelope glycoproteins 1 and 2 (gpE1 and gpE2), as well as priming MHC class-II-restricted CD4⁺ T helper (T_H) and MHC class-I-restricted CD8⁺ cytotoxic lymphocyte (CTL) responses to these and other viral proteins. Both types of T cell can secrete antiviral cytokines such as interferon- γ (IFN- γ), and CD8⁺ CTLs have the potential to kill infected cells.

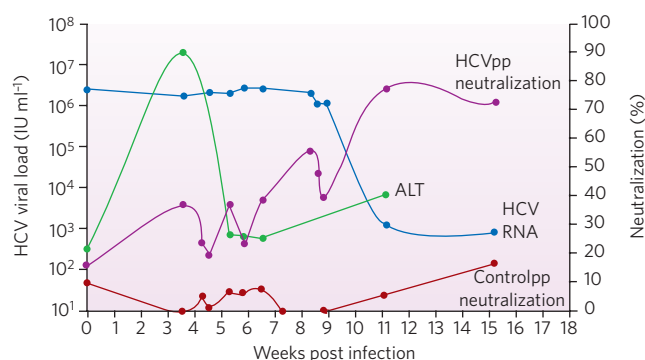


Figure 1 | Putative control of acute HCV viraemia by viral neutralizing antibodies. Association between circulating viral RNA load (blue) and antibodies that neutralize infectivity of HCV pseudoparticles (HCVpp; purple, expressed as percentage of neutralization). The percentage of neutralization of control pseudoparticles (Controlpp) is shown in red. Serum alanine aminotransferase (ALT) levels indicative of hepatitis are shown in green. Adapted from ref. 29.

It is difficult to prime CD8⁺ CTLs using polypeptide subunit vaccines, although certain adjuvants are capable of eliciting such responses^{13,14}. Various forms of plasmid DNA vaccine are also being explored to elicit HCV-specific humoral and cellular immune responses to encoded antigens which, by virtue of being newly synthesized in the cytosol of transfected cells, can be particularly effective at priming CD8⁺ CTLs. DNA vaccines also include immunostimulatory deoxycytosine-deoxyguanosine (CpG)-containing motifs capable of activating antigen-presenting dendritic cells. This would lead to stimulation of innate immune responses (such as the synthesis of type I interferons and natural killer (NK) cells) as well as adaptive B- and T-cell responses to vaccine antigens. Various live attenuated or defective viral or bacterial vectors expressing HCV genes are also being investigated because improved vaccine immunogenicity can result from more efficient expression and delivery of HCV antigens. This may include the targeting of antigen-presenting cells in some cases. The use of various

¹Chiron Corporation, 4560 Horton Street, Emeryville, California 94608, USA; ²National Institute of Molecular Genetics, INGM, Via Francesco Sforza 28, 20122, Milan, Italy.

prime/boost immunization modes and regimens are also being explored to optimize vaccine immunogenicity and potency.

In this review, we will summarize current knowledge regarding the correlates of immunity to HCV as well as the results of pre-clinical studies using vaccine candidates designed to recapitulate protective immunity. Emerging results from the chimpanzee challenge model suggest that successful vaccination against homologous and at least some heterologous HCV strains may be feasible, although the relative roles of humoral and cellular immunity in protection need to be better defined. The status and issues surrounding clinical development will be discussed as well as the rationale and prospect for immunotherapeutic vaccination strategies.

Correlates of immunity

Infected humans and chimpanzees who mount an early, multi-specific CD4⁺ T_H and CD8⁺ T-cell response to HCV proteins can eradicate the virus (see the review in this issue by Bowen and Walker, page 946, and refs 15–24). These activated T cells secrete pro-inflammatory cytokines (T_H1-type) such as IFN- γ , which is directly antiviral for HCV replicons in cell culture²⁵ and temporally associated with large reductions in viral load during acute infection²⁰. Specific CD8⁺ CTLs can kill HCV-infected cells, although large reductions in viral load during acute infection were not associated with an increase in acute hepatitis, suggesting that cytolytic activity may not be the main factor in viral control²⁰. Further work is required, however, to better define these cellular correlates of immunity. It is clear that these cellular immune responses to the virus can occur in the absence of antibody to gpE1 and gpE2 (ref. 19), indicating that such antibodies are not absolutely required for recovery from acute infection.

Until very recently^{10–12}, HCV has not been propagated efficiently in cell culture, meaning that a direct assay for viral-neutralizing antibody has not been available. However, the recent production of lentiviral/HCV pseudoparticles (HCVpp) bearing HCV envelope glycoproteins on the particle surface have been used to show that patients not only have antibodies that can neutralize the infectivity of such pseudoparticles but that such antibodies cross-neutralize pseudoparticles derived from many different HCV genotypes^{26–28}. This suggests that a broad cross-neutralizing antibody to HCV may exist and could be exploited in vaccine strategies. Furthermore, the recent application of these pseudoparticle infectivity assays to the investigation of immune correlates of protection are beginning to indicate that such ‘neutralizing’ antibodies, when present, may be associated with recovery from acute infection, at least in some cases²⁹ (for an example, see Fig. 1). The relative roles of humoral and cellular immunity in recovery remain unclear.

Studies in the past several years have helped to define the sophisticated battle initiated in the infected host. This RNA virus (which cannot integrate into the host genome) has evolved mechanisms to persist and to evade the host's innate and adaptive immune mechanisms (see the reviews in this issue by Gale and Foy, page 939, and Bowen and Walker, page 946). The virus inhibits the induction of type-1 interferons^{30,31}, inhibits NK cells^{32,33}, readily produces escape mutants to CTLs³⁴ and neutralizing antibodies directed to the

amino-terminal region of gpE2^{35–38}. HCV may also inhibit the binding of virion-neutralizing antibodies by masking with lipoproteins^{28,39}. Effector T cells specific to the virus also seem to be downregulated in some way as a consequence of persistent HCV infection^{40,41}. Other mechanisms of viral persistence are likely to emerge in the future. By contrast, if the host has the ability to elicit early and broad T_H1-type CD4⁺ and CD8⁺ T-cell responses to the virus^{15–24} and also has a NK receptor repertoire that facilitates innate immune clearance of virus⁴², then eradication of virus can occur. It is also possible that the presence of viral-neutralizing antibody may enhance this process²⁹.

Prevention strategies

Results from our recent studies have made us optimistic about successfully vaccinating against HCV. These studies involved the use of the recombinant HCV envelope glycoproteins gpE1 and gpE2 as vaccine antigens. Derived from mammalian cells, the two glycoproteins associate together to form a non-disulphide linked gpE1–gpE2 heterodimer that is thought to resemble the pre-virion envelope structure⁴³. When combined with oil/water-based adjuvants and used to vaccinate naive chimpanzees, this vaccine candidate elicits anti-envelope antibodies as well as T_H cell responses to gpE1 and gpE2. Our earlier work showed that when these vaccinated animals were challenged experimentally with homologous viral inocula, the highest responding animals (in terms of anti-gpE1/gpE2 antibody titres) were completely protected against infection⁴⁴. Using sensitive RT-PCR assays, no viraemia was detected in blood or liver samples at any time after challenge in these seemingly ‘sterilized’ animals. This apparent sterilizing immunity correlated directly with anti-gpE2 antibody titres that prevent the binding of gpE2 (or the virus itself) to CD81 (ref. 45), which has been shown to be an important receptor component for binding of infectious HCV^{10–12,46} and for cell entry of lentiviral/HCV pseudoparticles⁴⁷. Furthermore, although lower-responding animals became infected, the majority underwent an abortive acute infection that did not result in the persistently infected carrier state^{44,48} that in humans can be associated with chronic liver disease^{1,4}. Overall, these data showed that the carrier rate in vaccinees was significantly lower than in unimmunized controls^{44,48} (Table 1).

A crucial question that remained was whether the vaccine derived from strain HCV-1 would protect against heterologous strains of the virus. Recently, we have challenged nine chimpanzee vaccinees with the HCV-H strain that, like the vaccine strain HCV-1, is of the 1a genotype that predominates in the United States³. Although none of the vaccinated animals was protected against acute infection, all but one vaccinee resolved the acute infection and failed to progress to the carrier state⁴⁹ (as demonstrated by the persistent absence of detectable viraemia in follow-up blood samples using sensitive RT-PCR assays). By contrast, the majority of control animals became carriers when challenged with HCV-H, indicating that the vaccine significantly reduced chronic, persistent infection⁴⁹ (Table 1). Although the viral challenge doses were small (10–100 chimpanzee infectious doses₅₀ (CID₅₀)), such doses are considered to be within the same range as those transmitted

Table 1 | Summary of outcome of experimental challenges of chimpanzees immunized with recombinant gpE1/gpE2

	Group	Total	Acute infections	Chronic infections	
Homologous HCV-1 challenges	Adjuvanted gpE1/gpE2	12	7	2 (17%)	$P = 0.03$
	Unimmunized controls	10	10	7 (70%)	
Heterologous HCV-H challenges	Adjuvanted gpE1/gpE2	9	9	1 (11%)	$P = 0.04$
	Unimmunized controls	14	14	8 (57%)	
Totals	Adjuvanted gpE1/gpE2	21	16	3 (14%)	$P = 0.002$
	Unimmunized controls	24	24	15 (63%)	

Typically, animals were immunized with 30–80 μ g gpE1/gpE2 in oil/water adjuvants on months 0, 1 and 7 approximately, followed by intravenous challenge on month 8 with 10–100 infectious doses of HCV-1 or HCV-H. Circulating levels of viraemia were measured using RT-PCR assays for HCV genomic RNA for at least one year post challenge^{44,48,49}. P values refer to chronic carrier rates between controls and vaccinees.

Table 2 | Prophylactic HCV vaccine candidates

Vaccine	Potency	Stage
Recombinant gpE1/gpE2 in oil/water adjuvants ^{44,48,49}	Protects chimps against chronic infection	Phase 1 clinical trials
DNA prime and protein boost (using C, gpE1, gpE2 and NS3) ⁵¹	Protection or amelioration in chimp challenge model	Pre-clinical
Recombinant VLPs containing C, gpE1 and gpE2 ⁶¹	Highly immunogenic in mice and baboons	Pre-clinical
Recombinant gpE1 in alum ⁷¹	Primes humoral and cellular immune responses in humans	Phase 1/2 clinical trials
Modified vaccinia ankara expressing gpE1/gpE2 ⁵⁸	Induces T _H 1 response in HLA A2.1 mice	Pre-clinical
Semliki forest virus expressing NS3 ⁵⁶	Induces NS3-specific CTLs in mice	Pre-clinical
DNA encoding gpE1/gpE2 in poly-lactide-co-glycolide particles ⁶⁰	Substantial increase in anti-gpE1/gpE2 titre in mice compared with naked DNA	Pre-clinical
Defective ovine adenovirus expressing NS3 ⁵⁹	Strong T _H 1 cellular response in mice	Pre-clinical
DNA prime and canary pox boost (encoding all HCV genes) ⁵⁵	Broad T _H 1 cellular immune responses in mice	Pre-clinical
Vaccinia virus expressing all HCV genes (A. Prince; personal communication)	Chimpanzee challenge studies in progress	Pre-clinical
Defective alphaviral particles expressing gpE1/gpE2 and NS genes ⁵⁷	Mouse studies in progress	Pre-clinical

in many community-acquired HCV infections because the infectivity titre of most carriers is known to be low⁵⁰. These pre-clinical data (and supporting data from other small studies exploring various gpE1/gpE2 vaccine formulations^{51–53}) support the initiation of a clinical prophylactic programme using adjuvanted gpE1/gpE2 that is currently in phase 1 testing.

Many studies correlate recovery from acute HCV infection with cellular immune responses to the virus, and so other relevant strategies for developing a vaccine will involve eliciting a broad cellular immune response to the virus or, preferably, both a humoral (anti-gpE1/gpE2) and a cellular immune response. One small study using the chimpanzee model investigated the use of a vaccination regimen employing multiple immunizations with plasmid DNA encoding the nucleocapsid (C), gpE1, gpE2 and nonstructural protein 3 (NS3) domains followed by multiple boosting with an adjuvanted mixture of recombinant C, gpE1, gpE2 and NS3 proteins⁵¹. Following challenge with a heterologous strain (which causes chronic, persistent infection in the large majority of control animals), one vaccinee experienced an ameliorated and abortive acute infection that did not progress to the carrier state, whereas the other vaccinee developed chronic infection, albeit ameliorated in terms of viral load and level of hepatitis. These data provide additional support for the feasibility of successful vaccination against HCV but also suggest that further optimization of vaccine immunogenicity is required.

Surprisingly, eliciting broad CD4⁺ and CD8⁺ T-cell responses to the virus in the absence of any antibody responses to the envelope glycoproteins (using an ISCOMATRIX[®]-adjuvanted^{13,14} NS3-4-5-Core polyprotein derived from strain HCV-1) failed to prevent chronic, persistent infection following challenge with the heterologous HCV-H strain in five out of five chimpanzee vaccinees tested, despite observing a substantial amelioration in acute viraemia and hepatitis (M.H., unpublished data). This result may be caused by insufficient priming of cellular immune responses by the vaccine regimen or protocol because recovery from acute infection has been linked with cellular immune responses to the virus in the absence of anti-envelope antibody responses¹⁹. This result also suggests that vaccine formulations capable of priming both anti-envelope neutralizing antibody and broad cellular immune responses to the virus may be more effective. Considering that this is the only vaccine formulation of several tested by us that failed to result in prevention of

HCV chronicity in at least some animals, this result also emphasizes the importance of using the chimpanzee challenge model before proceeding to clinical testing.

Other approaches to HCV vaccination (summarized in Table 2), in common with those used in vaccine research for other persistent pathogens like HIV and malaria, include the use of various defective or attenuated viral vectors to enhance priming of humoral and cellular immune responses to multiple HCV gene products expressed by the vector. The use of adenoviral⁵⁴, avipox⁵⁵, alphaviral^{56,57} and vaccinia⁵⁸ viral vectors, among others, are all being explored in various animal models including the chimpanzee challenge model. These approaches offer the potential of improved immunogenicity as a result of enhanced gene delivery and expression. Some of these vectors also infect and/or activate antigen-presenting dendritic cells, thus enhancing antigen presentation and stimulating innate immune responses that, in turn, lead to enhancement of adaptive immune responses to the encoded vaccine antigens. However, challenges for these approaches include the problem of pre-existing immunity to some of these vectors in the human population, thus limiting potency. Repeated vaccination to boost initial immune responses can also be limited by vector-elicited immunity. To overcome this obstacle, priming of the immune response with DNA vaccines followed by boosting with recombinant viral vectors is being employed as well as prime/boost regimens using different recombinant vectors for each immunization. One promising approach is the use of defective alphaviral delivery vectors that infect professional antigen-presenting dendritic cells, activate innate immunity as well as adaptive cellular and humoral immune responses to encoded vaccine antigens and which can be used repeatedly to boost immune responses in mice⁵⁷. A defective ovine adenovirus vector⁵⁹ may also be useful in this regard. Other promising approaches being explored include the use of DNA microparticles⁶⁰, which can significantly enhance the potency of DNA vaccines and HCV viral-like particles produced in insect cells that have an inherently strong immunogenicity⁶¹.

Apart from optimizing vaccine formulations to maximize humoral and cellular immune responses, future issues include expanding the range and level of cross-protection afforded by the vaccine. This will require more extensive analyses into the nature and range of cross-neutralizing antibody and cross-protective cellular

Table 3 | HCV immunotherapeutic vaccine candidates

Vaccine	Potency	Stage
Alum-adsorbed gpE1 glycoprotein ⁶⁸	Boosts humoral and cellular immune responses to gpE1 in HCV patients. May ameliorate hepatitis	Phase 1/2 patient trials
Adjuvanted peptide cocktails (A. von Gabain, Intercell, personal communication)	Designed to boost CD4 ⁺ and CD8 ⁺ responses to conserved T-cell epitopes using novel adjuvant	Phase 2 patient trials
Oil/water-adsorbed gpE1/gpE2 proteins ^{44,48,49}	Prophylactic efficacy in chimpanzees. Boosts anti-gpE1/gpE2 antibody titres in chronically infected HCV chimpanzees	Phase 1b patient trials
ISCOMATRIX®-adjuvanted ^{13,14} Core protein	Primes T _H 1-type CD4 ⁺ and CD8 ⁺ CTL responses in macaques and uninfected humans to conserved epitopes within Core antigen	Phase 1b patient trials
ISCOMATRIX®-adjuvanted ^{13,14} NS3-NS4-NS5-C polyprotein (M.H., unpublished data)	Primes broad T _H 1-type CD4 ⁺ and CD8 ⁺ CTL responses in chimpanzees which when challenged with heterologous HCV have reduced viraemia and hepatitis relative to controls	Pre-clinical
Prime with recombinant adenovirus and boost with electroporated DNA (expressing HCV NS 3, 4 and 5) (A. Nicosia, personal communication)	Primes broad CD4 ⁺ and CD8 ⁺ T cells. Chimpanzee challenge studies in progress	Pre-clinical
Heat-killed yeast expressing C and NS3 ⁷²	Primes specific CD4 ⁺ and CD8 ⁺ T cells in mice	Pre-clinical

immune responses and will probably involve the definition of cocktails of immunogens derived from various HCV genotypes to obtain an effective global vaccine formulation.

Proving the efficacy of the vaccine in humans is a significant challenge because accessing groups at high risk of HCV infection is no longer a simple task. With the near elimination of post-transfusion hepatitis C by donor screening, other high-risk groups suitable for efficacy testing have inherent difficulties such as lack of compliance (for example, intravenous drug users), low incidence of infection (for example, in health-care workers and paramedics), lack of supporting infrastructure (for example, in many developing countries where incidence of infection is high) and ethical issues (for example, in prisoner populations where prevalence and incidence of infection are both high). However, some of these cohorts have been used successfully in the past (for testing hepatitis B vaccines) and so these obstacles should not be insurmountable. If a vaccine is successfully developed, an important cause of global morbidity and mortality will be controlled and, even in countries with a relatively low incidence of infection, the vaccine will be reasonably cost-effective when used in the general population⁶².

Potential for therapeutic HCV vaccination

The current standard-of-care therapy for chronically infected HCV patients is a combination of pegylated IFN- α and ribavirin, which is costly, lengthy (6–12 months), associated with significant side effects and results in sustained viral response in only ~50% of patients. In patients infected with genotype 1, the most common form, response rates are even lower⁶⁵. With an estimated 170 million HCV carriers worldwide, it is clearly important to develop better therapeutic options. With our increasing knowledge of the virus-encoded enzymes and genetic elements vital to the life-cycle of HCV, much attention is now being focused on the development of HCV protease, replicase, helicase, antisense, silencing RNA and other specific inhibitors. However, preliminary data have directly linked responses to IFN- α and ribavirin with pretreatment titres of viral antibodies⁶⁴ (presumed to be against the envelope glycoproteins), peripheral T_H cell responses to the HCV core and other antigens⁶⁵, as well as to intrahepatic CD8⁺ CTL responses to the virus⁶⁶. Total pretreatment CD8⁺ T-cell counts in the liver have also been correlated with sustained responses to standard-of-care therapy⁶⁷. Therefore, it may be possible to boost such immune responses in patients by appropriate vaccination and thereby improve the response rate to the standard-of-care therapy. Such immunotherapy may also help control the emergence of escape mutants that would

be predicted to arise from any future use of HCV protease or replicase inhibitors, for example, given the extreme fluidity and heterogeneity of the HCV genome³.

Many therapeutic vaccine trials are planned or are already in progress and use diverse delivery methods and formulations (summarized in Table 3) but little information is available about their efficacy at present. What is known, however, is that use of an alum-adsorbed recombinant gpE1 antigen was able to boost humoral and cellular immune responses to gpE1 in viraemic patients, providing encouragement that vaccination can increase immune responses in pre-existing carriers⁶⁸. It remains to be seen whether boosting viral-neutralizing antibody titres or broad CD4⁺ T_H responses or broad CD8⁺ T-cell responses will have the greatest impact on reducing viral load and in the response to antiviral therapy. But, as may be the case for optimal prophylaxis, boosting all of these immune responses may be ideal for immunotherapy.

HCV tries to counter innate immunity by inhibiting the induction of type-I interferons (IFN- α/β)^{30,31} and downregulating NK-cell activity^{32,33}. Therefore, therapeutic vaccine formulations could benefit by inclusion of molecules capable of triggering innate immune responses. Such molecules include oligonucleotides containing CpG motifs that trigger Toll-like receptor 9 within dendritic cells and that also enhance adaptive immune responses to vaccine antigens⁶⁹. If successful, vaccination for the treatment of chronic hepatitis C would be one of the first demonstrations of immunotherapeutic intervention in chronic viral infections, although, very recently, such an approach has been used successfully to inhibit the age-related emergence of herpes zoster infections and disease in carriers⁷⁰.

Future directions

In the future, it will be important to use the chimpanzee model to further define correlates of protection, duration of vaccine-mediated protection, the extent of cross-protection against diverse genotypes and mechanisms of chronicity and to determine optimal vaccine formulations for prophylactic and immunotherapeutic efficacy. In addition, human cohorts at high risk of infection need to be identified and characterized for efficacy trials. The huge burden of chronically infected HCV patients facilitates the testing of various immunotherapeutic vaccine formulations that, most probably, will be especially useful when used as adjunct therapy with antiviral drugs, including pegylated IFN- α and ribavirin as well as the new class of HCV drugs currently under development that inhibit viral enzymes and other elements crucial to the viral life-cycle. It will also be important to under-

stand the mechanisms involved in immune dysfunction and evasion during chronic HCV infections so as to facilitate the design of further immunotherapies. ■

- Alter, H. J. & Seeff, L. B. Recovery, persistence, and sequelae in hepatitis C virus infection: a perspective on long-term outcome. *Semin. Liver Dis.* **20**, 17–35 (2000).
- Lai, M. E. et al. Hepatitis C virus in multiple episodes of acute hepatitis in polytransfused thalassaemic children. *Lancet* **343**, 388–390 (1994).
- Simmonds, P. Genetic diversity and evolution of hepatitis C virus—15 years on. *J. Gen. Virol.* **85**, 3173–3188 (2004).
- Seeff, L. B. Natural history of chronic hepatitis C. *Hepatology* **36**, S35–S46 (2002).
- Mehta, S. H. et al. Protection against persistence of hepatitis C. *Lancet* **359**, 1478–1483 (2002).
- Weiner, A. J. et al. Intrahepatic genetic inoculation of hepatitis C virus RNA confers cross-protective immunity. *J. Virol.* **75**, 7142–7148 (2001).
- Bassett, S. E. et al. Protective immune response to hepatitis C virus in chimpanzees rechallenged following clearance of primary infection. *Hepatology* **33**, 1479–1487 (2001).
- Lanford, R. E. et al. Cross-genotype immunity to hepatitis C virus. *J. Virol.* **78**, 1575–1581 (2004).
- Farci, P. et al. Lack of protective immunity against reinfection with hepatitis C virus. *Science* **258**, 135–140 (1992).
- Wakita, T. et al. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Med.* **11**, 791–796 (2005).
- Zhong, J. et al. Robust hepatitis C virus infection *in vitro*. *Proc. Natl Acad. Sci. USA* **102**, 9294–9299 (2005).
- Lindenbach, B. D. et al. Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623–626 (2005).
- Polakos, N. K. et al. Characterization of hepatitis C virus core-specific immune responses primed in rhesus macaques by a non-classical ISCOM vaccine. *J. Immunol.* **166**, 3589–3598 (2001).
- Pearse, M. J. & Drane, D. ISCOMATRIX adjuvant for antigen delivery. *Adv. Drug Deliv. Rev.* **57**, 465–474 (2005).
- Diepolder, H. M. et al. Possible mechanism involving T-lymphocyte response to non-structural protein 3 in viral clearance in acute hepatitis C virus infection. *Lancet* **346**, 1006–1007 (1995).
- Missale, G. et al. Different clinical behaviors of acute hepatitis C virus infection are associated with different vigor of the anti-viral cell-mediated immune response. *J. Clin. Invest.* **98**, 706–714 (1996).
- Tsai, S. L., Liaw, Y. F., Chen, M. H., Huang, C. Y. & Kuo, G. C. Detection of type 2-like T-helper cells in hepatitis C virus infection: implications for hepatitis C virus chronicity. *Hepatology* **25**, 449–458 (1997).
- Gerlach, J. T. et al. Recurrence of hepatitis C virus after loss of virus-specific CD4⁺ T-cell response in acute hepatitis C. *Gastroenterology* **117**, 933–941 (1999).
- Cooper, S. et al. Analysis of a successful immune response against hepatitis C virus. *Immunity* **10**, 439–449 (1999).
- Thimme, R., Oldach, D., Chang, K. M., Steiger, C., Ray, S. C. & Chisari, F. V. Determinants of viral clearance and persistence during acute hepatitis C virus infection. *J. Exp. Med.* **194**, 1395–1406 (2001).
- Grakoui, A. et al. HCV persistence and immune evasion in the absence of memory T cell help. *Science* **302**, 659–662 (2003).
- Shoukry, N. H. et al. Memory CD8⁺ T cells are required for protection from persistent hepatitis C virus infection. *J. Exp. Med.* **197**, 1645–1655 (2003).
- Lechner, F. et al. Analysis of successful immune responses in persons infected with hepatitis C virus. *J. Exp. Med.* **191**, 1499–1512 (2000).
- Lechner, F. et al. CD8⁺ T lymphocyte responses are induced during acute hepatitis C virus infection but are not sustained. *Eur. J. Immunol.* **30**, 2479–2487 (2000).
- Frese, M. et al. Interferon-gamma inhibits replication of subgenomic and genomic hepatitis C virus RNAs. *Hepatology* **35**, 694–703 (2002).
- Bartosch, B. et al. *In vitro* assay for neutralizing antibody to hepatitis C virus: evidence for broadly conserved neutralization epitopes. *Proc. Natl Acad. Sci. USA* **100**, 14199–14204 (2003).
- Logvinoff, C. et al. Neutralizing antibody response during acute and chronic hepatitis C virus infection. *Proc. Natl Acad. Sci. USA* **101**, 10149–10154 (2004).
- Meunier, J. C. et al. Evidence for cross-genotype neutralization of hepatitis C virus pseudo-particles and enhancement of infectivity by apolipoprotein C1. *Proc. Natl Acad. Sci. USA* **102**, 4560–4565 (2005).
- Lavillette, D. et al. Human serum facilitates hepatitis C virus infection, and neutralizing responses inversely correlate with viral replication kinetics at the acute phase of hepatitis C virus infection. *J. Virol.* **79**, 6023–6034 (2005).
- Foy, E. et al. Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-1 signaling. *Proc. Natl Acad. Sci. USA* **102**, 2986–2991 (2005).
- Li, K. et al. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc. Natl Acad. Sci. USA* **102**, 2992–2997 (2005).
- Crotta, S. et al. Inhibition of natural killer cells through engagement of CD81 by the major hepatitis C virus envelope protein. *J. Exp. Med.* **195**, 35–41 (2002).
- Tseng, C. T. & Klimpel, G. R. Binding of the hepatitis C virus envelope protein E2 to CD81 inhibits natural killer cell functions. *J. Exp. Med.* **195**, 43–49 (2002).
- Erickson, A. L. et al. The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes. *Immunity* **15**, 883–895 (2001).
- Weiner, A. J. et al. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc. Natl Acad. Sci. USA* **89**, 3468–3472 (1992).
- Kato, N. et al. Humoral immune response to hypervariable region 1 of the putative envelope glycoprotein (gp70) of hepatitis C virus. *J. Virol.* **67**, 3923–3930 (1993).
- Farci, P. et al. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339–344 (2000).
- Farci, P. et al. Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *Proc. Natl Acad. Sci. USA* **93**, 15394–15399 (1996).
- Thomssen, R., Bonk, S., Propfe, C., Heermann, K. H., Kochel, H. G. & Uy, A. Association of hepatitis C virus in human sera with beta-lipoprotein. *Med. Microbiol. Immunol.* **181**, 293–300 (1992).
- Wedemeyer, H. et al. Impaired effector function of hepatitis C virus-specific CD8⁺ T cells in chronic hepatitis C virus infection. *J. Immunol.* **169**, 3447–3458 (2002).
- Semmo, N. et al. Preferential loss of IL-2-secreting CD4⁺ T helper cells in chronic HCV infection. *Hepatology* **41**, 1019–1028 (2005).
- Khakoo, S. I. et al. HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* **305**, 872–874 (2004).
- Ralston, R. et al. Characterization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia viruses. *J. Virol.* **67**, 6753–6761 (1993).
- Choo, Q. L. et al. Vaccination of chimpanzees against infection by the hepatitis C virus. *Proc. Natl Acad. Sci. USA* **91**, 1294–1298 (1994).
- Rosa, D. et al. A quantitative test to estimate neutralizing antibodies to the hepatitis C virus: cytofluorimetric assessment of envelope glycoprotein 2 binding to target cells. *Proc. Natl Acad. Sci. USA* **93**, 1759–1763 (1996).
- Pileri, P. et al. Binding of hepatitis C virus to CD81. *Science* **282**, 938–941 (1998).
- McKeating, J. A. et al. Diverse hepatitis C virus glycoproteins mediate viral infection in a CD81-dependent manner. *J. Virol.* **78**, 8496–8505 (2004).
- Houghton, M. et al. *In Viral Hepatitis and Liver Disease* (eds Rizzetto, M., Purcell, R. H., Gerin, J. L., Verme, G.) pp 656–659 (Edizioni Minerva Medica, 1997).
- Coates, S. et al. in *Proceedings of the 11th International Symposium on Viral Hepatitis and Liver Disease* (eds Jilbert, A. R., Grgacic, E. V. L., Vickery, K., Burrell, C. J., Cossart, Y. E.) pp 118–123 (Australian Center for Hepatitis Virology, 2005).
- Prince, A. M. Reliability of chimpanzee model for non-A, non-B hepatitis. *Lancet* **ii**, 1134 (1985).
- Rollier, C. et al. Control of heterologous hepatitis C virus infection in chimpanzees is associated with the quality of vaccine-induced peripheral T-helper immune response. *J. Virol.* **78**, 187–196 (2004).
- Puig, M., Major, M. E., Mihalik, K. & Feinstone, S. M. Immunization of chimpanzees with an envelope protein-based vaccine enhances specific humoral and cellular immune responses that delay hepatitis C virus infection. *Vaccine* **22**, 991–1000 (2004).
- Forns, X. et al. Vaccination of chimpanzees with plasmid DNA encoding the hepatitis C virus (HCV) envelope E2 protein modified the infection after challenge with homologous monoclonal HCV. *Hepatology* **32**, 618–625 (2000).
- Catalucci, D., Sporeno, E., Cirillo, A., Ciliberto, G., Nicosia, A. & Colloca, S. An adenovirus type 5 (Ad5) amplicon-based packaging cell line for production of high-capacity helper-independent deltaE1-E2-E3-E4 Ad5 vectors. *J. Virol.* **79**, 6400–6409 (2005).
- Pancholi, P., Perkus, M., Tricoche, N., Liu, Q. & Prince, A. M. DNA immunization with hepatitis C virus (HCV) polycistronic genes or immunization by HCV DNA priming-recombinant canarypox virus boosting induces immune responses and protection from recombinant HCV-vaccinia virus infection in HLA-A2.1-transgenic mice. *J. Virol.* **77**, 382–390 (2003).
- Brinster, C. et al. Hepatitis C virus non-structural protein 3-specific cellular immune responses following single or combined immunization with DNA or recombinant Semliki Forest virus particles. *J. Gen. Virol.* **83**, 369–381 (2002).
- Perri, S. et al. An alphavirus replicon particle chimera derived from venezuelan equine encephalitis and sindbis viruses is a potent gene-based vaccine delivery vector. *J. Virol.* **77**, 10394–10403 (2003).
- Abraham, J. D. et al. Comparative immunogenicity analysis of modified vaccinia Ankara vectors expressing native or modified forms of hepatitis C virus E1 and E2 glycoproteins. *Vaccine* **22**, 3917–3928 (2004).
- Wuest, T., Both, G. W., Prince, A. M., Hofmann, C. & Loser, P. Recombinant ovine atadenovirus induces a strong and sustained T cell response against the hepatitis C virus NS3 antigen in mice. *Vaccine* **22**, 2717–2721 (2004).
- O'Hagan, D. T. et al. Cationic microparticles are a potent delivery system for a HCV DNA vaccine. *Vaccine* **23**, 672–680 (2004).
- Jeong, S. H. et al. Immunization with hepatitis C virus-like particles induces humoral and cellular immune responses in nonhuman primates. *J. Virol.* **78**, 6995–7003 (2004).
- Krahn, M. D. et al. Potential cost-effectiveness of a preventive hepatitis C vaccine in high risk and average risk populations in Canada. *Vaccine* **23**, 1549–1558 (2005).
- Saadeh, S. & Davis, G. L. The evolving treatment of chronic hepatitis C: where we stand a decade out. *Cleveland Clin. J. Med.* **71** (Suppl. 3), S3–S7 (2004).
- Baumert, T. F. et al. Antibodies against hepatitis C virus-like particles and viral clearance in acute and chronic hepatitis C. *Hepatology* **32**, 610–617 (2000).
- Cramp, M. E., Rossol, S., Chokshi, S., Carucci, P., Williams, R. & Naoumov, N. V. Hepatitis C virus-specific T-cell reactivity during interferon and ribavirin treatment in chronic hepatitis C. *Gastroenterology* **118**, 346–355 (2000).
- Nelson, D. R., Marousis, C. G., Ohno, T., Davis, G. L. & Lau, J. Y. Intrahepatic hepatitis C virus-specific cytotoxic T lymphocyte activity and response to interferon alpha therapy in chronic hepatitis C. *Hepatology* **28**, 225–230 (1998).
- Vrolijk, J. M. et al. Pretreatment intrahepatic CD8⁺ cell count correlates with virological

- response to antiviral therapy in chronic hepatitis C virus infection. *J. Infect. Dis.* **188**, 1528–1532 (2003).
68. Nevens, F. *et al.* A pilot study of therapeutic vaccination with envelope protein E1 in 35 patients with chronic hepatitis C. *Hepatology* **38**, 1289–1296 (2003).
69. Abel, K. *et al.* Deoxycytidyl-deoxyguanosine oligonucleotide classes A, B, and C induce distinct cytokine gene expression patterns in Rhesus monkey peripheral blood mononuclear cells and distinct alpha interferon responses in TLR9-expressing Rhesus monkey plasmacytoid dendritic cells. *Clin. Diagn. Lab. Immunol.* **12**, 606–621 (2005).
70. Oxman, M. N. *et al.* A vaccine to prevent herpes zoster and postherpetic neuralgia in older adults. *N. Engl. J. Med.* **352**, 2271–2284 (2005).
71. Leroux-Roels, G. *et al.* A candidate vaccine based on the hepatitis C E1 protein: tolerability and immunogenicity in healthy volunteers. *Vaccine* **22**, 3080–3086 (2004).
72. Franzusoff, A., Duke, R. C., King, T. H., Lu, Y., & Rodell, T. C. Yeasts encoding tumour antigens in cancer immunotherapy. *Exp. Opin. Biol. Ther.* **5**, 565–575 (2005).

Author information Reprints and permissions information is available at npg.nature.com/preprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for material should be addressed to M.H. (michael_houghton@chiron.com).

Mechanism of action of interferon and ribavirin in treatment of hepatitis C

Jordan J. Feld¹ and Jay H. Hoofnagle²

Since the identification of the hepatitis C virus, great strides have been made in the development of an antiviral therapy. As a crucial mediator of the innate antiviral immune response, interferon- α (IFN- α) was a natural choice for treatment. Whereas treatment with IFN- α alone achieved only modest success, the addition of the broad-spectrum antiviral agent ribavirin greatly improved responses. However, half of the infected individuals with chronic disease do not achieve sustained clearance of hepatitis C virus. To optimize current therapeutic strategies and to develop new therapies, a better understanding of the mechanism of action of IFN and ribavirin will be essential.

Interferon- α (IFN- α) was first shown to have beneficial effects in patients with chronic hepatitis C in 1986, well before the identification of the hepatitis C virus (HCV)¹. Once HCV was identified and tests developed for its detection in serum, the basis for IFN activity became clear. IFN- α therapy led to a rapid decline in HCV-RNA levels in serum, and long-term responses were marked by sustained loss of HCV RNA from the serum and liver and resolution of the chronic infection².

Unfortunately, therapy with IFN- α alone (IFN- α monotherapy) had only limited success. A 6-month course led to sustained response rates of 6–12%, and extending treatment to 12 months raised this rate to only 16–20% (ref. 3). A major advance came with the addition of the broad-spectrum antiviral agent ribavirin to IFN- α treatment, which more than doubled the sustained response rate to 35–40% (ref. 4). Further improvement has recently been achieved by the development of pegylated interferon, in which a large molecule of poly(ethylene glycol) (PEG) is covalently attached to recombinant IFN- α , resulting in an active molecule with a longer half-life, better pharmacokinetic profile and better rate of virological response^{5–8}. The combination of pegylated interferon with ribavirin yields sustained response rates of 54–56% (refs 9–11). These results are heartening, but they also mean that 40–50% of patients do not have lasting improvement with treat-

ment. Furthermore, combination therapy is expensive, associated with frequent and troublesome side effects, and contraindicated in many patients. The response rates of 54–56% apply to selected populations without the co-morbidities that often accompany hepatitis C¹². Clearly, more-effective and better-tolerated therapies for hepatitis C are needed.

Unfortunately, the development of new, potent, specific agents against HCV has been difficult, and newer agents are likely to be used in combination with IFN- α (see the review in this issue by De Francesco and Migliaccio, page 953). For this reason, pegylated interferon and ribavirin are likely to remain the cornerstones of therapy for hepatitis C for the near future. Improvements in dosing, dose regimen and support of patients receiving pegylated interferon and ribavirin might materially improve response rates but more substantial improvements will come only with a better understanding of the mechanisms of action of these agents.

Pattern of virological responses to therapy

Responses to antiviral therapy of hepatitis C are grouped into three general patterns (Fig. 1): sustained virological response (SVR); end-of-treatment response and relapse; and non-response. The patterns have different implications. An SVR is defined as the loss of detectable HCV RNA during treatment and its continued absence for at least 6 months after stopping therapy. Several studies of long-term follow-up on patients who achieve an SVR demonstrate that this response is durable in over 95% of patients^{2,13}. Furthermore, liver histology improves, with resolution of inflammation and regression of fibrosis following viral clearance.

A transient response with relapse occurs in 10–25% of patients with optimal regimens. These patients show little evidence of a long-term benefit. The cause of relapse is not well understood but the rate correlates with shorter treatment, with inadequate or absent doses of ribavirin and, to a lesser extent, with the degree of hepatic fibrosis and cirrhosis. Retreatment of patients with relapse sometimes results in an SVR but usually only when a longer course or higher doses are used¹⁴. Finally, non-response to treatment occurs in about one-third of patients with chronic hepatitis C. These patients never become HCV-RNA negative, although titres might fall during treatment (Fig. 1). Elucidation of the causes for non-response to IFN- α therapy is a major challenge to research on HCV.

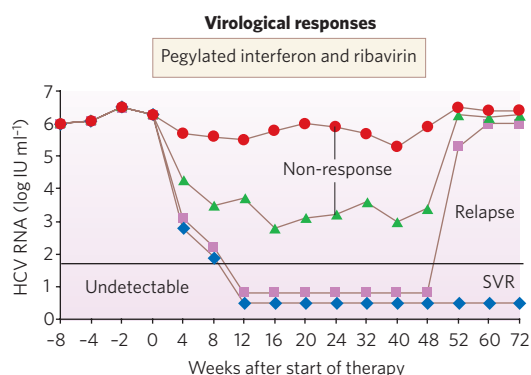


Figure 1 | Virological responses to hepatitis C therapy. Different patterns of viral response during interferon- α -based therapy of chronic hepatitis C. SVR, sustained virological response.

¹Liver Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases, Building 31, Room 9A27. ²31 Center Drive, National Institutes of Health, Bethesda, Maryland 20892, USA.

Responses to treatment are also characterized by different viral kinetic profiles. Virological responders typically have a very rapid initial decrease in viral level, followed by a second, slower phase of decline until undetectable levels of circulating virus are achieved¹⁵ (Fig. 2). The initial decrease (phase one) is believed to reflect the efficiency of suppression of replication and is calculated as a percentage decline in HCV RNA levels within the first 1–2 days after the initial injection of IFN. The first-phase decrease averages 90–99% (one to two log decline) but varies considerably, from 99.9% (> 3 logs) to 0%. The second-phase response is believed to be due to clearance of virus-infected cells (by cell death or by eradication of viral replication in the cell) and is calculated from the rate of decline in HCV RNA levels following the first-phase response. Importantly, with thrice-weekly treatment with standard IFN- α and weekly treatment with pegylated interferon, the classical viral kinetic response occurs in only a proportion of patients. Others exhibit paradoxical responses (with rebound or multiple phases of response) and some patients have no response at all¹⁶. Patients in the last category (null or flat response) are almost always non-responders. Why this occurs or what block in the mechanism of action of IFN and ribavirin accounts for this phenomenon is not known.

Determinants of response to therapy

The currently recommended therapy for chronic hepatitis C is a combination of pegylated interferon and ribavirin for 24 or 48 weeks³. Overall sustained responses occur in about one-half of patients but the likelihood of response varies greatly, depending on viral and host characteristics (Box 1), especially the viral genotype. SVR rates range from 42% to 46% in patients with genotype 1, which accounts for ~70% of cases in the USA^{9–11,17}. By contrast, SVR rates for patients with the less common genotypes 2 and 3 are 76–80%. Furthermore, patients with genotypes 2 and 3 can be treated with a shorter course of therapy and with lower doses of ribavirin with no sacrifice in response rate¹¹.

Serum concentration of HCV RNA at the time of initiation of antiviral therapy is another determinant of outcome in patients with genotype-1 infection. Sustained response rates are consistently higher in patients with low baseline HCV RNA levels (usually defined as < 800,000 IU ml⁻¹). Host factors also affect the chance of SVR, albeit less so than the genotype does. These include age, race, gender, obesity and degree of hepatic fibrosis. Among these factors, racial differences in response rates are the most striking. African Americans with chronic hepatitis C have responses that are one-half to one-third those in Caucasians¹⁸. The reasons for the racial differences in response rates to pegylated interferon and ribavirin therapy for hepatitis C are not known.

An important clinical observation about antiviral therapy of hepatitis C is the high rate of response in patients with acute hepatitis C¹⁹. Acute infection with HCV is marked by a high rate of viral persistence, with chronic infection evolving in 50–80% of patients. IFN- α therapy during acute hepatitis C reduces the chronicity rate to 10% or lower. Strikingly, almost all patients with acute hepatitis C (regardless of genotype or initial viral load) rapidly become HCV RNA negative on therapy^{19,20}. These findings suggest that non-response to IFN- α might be acquired during the establishment of chronic infection. Clinical investigations of patients with non-response to IFN-based therapy have focused on various issues: viral strain, sequence and quasispecies diversity; pharmacokinetic profiles (IFN and ribavirin dose and drug levels); disease characteristics (severity and activity of hepatitis); the effects of co-morbidities (obesity, diabetes, renal disease and immunodeficiency); problems in IFN cell signalling and actions; and potentially modifiable environmental factors (alcohol, smoking and adjunctive medications).

Mechanism of action of therapeutic IFN

The type-1 IFNs include interferons- α , β , ω and λ , all of which play a crucial role in the innate antiviral immune response^{21,22}. There are at least 14 IFN- α genes, but only one gene copy each for IFN- β and IFN- ω , and three copies for IFN- λ . All type-1 IFNs have antiviral, antiproliferative and immunomodulatory activities, but their relative potencies differ.

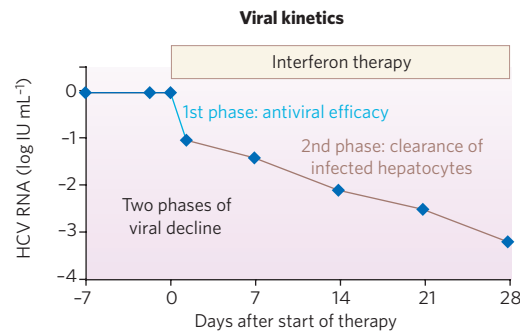


Figure 2 | Pattern of decrease in HCV RNA levels during interferon therapy of chronic hepatitis C. This shows a rapid initial first-phase decline in HCV RNA levels that occurs during the first 1–2 days of treatment, followed by a more gradual second-phase decline during the subsequent weeks of treatment.

Most forms of type-1 IFN have activity against HCV, yet few have been evaluated clinically²³. The current commercially available forms of IFN- α used for hepatitis C (α 2a, α 2b and consensus IFN) have somewhat different potencies *in vitro* but appear to yield similar response rates in treated patients. The type-1 IFNs might have similar clinical activities because they share, at least in part, cell-surface receptors and intracellular pathways of action. IFN- γ is classified as a type-1 IFN, in that it has antiviral activity *in vitro*. Interestingly, IFN- γ has activity against HCV in cell-culture systems²⁴ but not effect on HCV RNA levels in humans²⁵.

IFN- α has potent antiviral activity but does not act directly on the virus or replication complex. Rather, it acts by inducing IFN-stimulated genes (ISGs), which establish a non-virus-specific antiviral state within the cell^{21,22}. In brief, circulating IFN- α binds to IFN cell-surface-receptor subunits, leading to their dimerization and the activation of the receptor-associated Janus-activated kinase 1 (Jak1) and tyrosine kinase 2 (Tyk2)^{26,27}. The activated kinases phosphorylate the signal transducer and activator of transcription proteins 1 and 2 (STAT1 and STAT2). The activated STAT1/2 complex is then translocated to the cell nucleus, where it combines with IFN-regulatory factor 9 (IRF-9) to form a complex that binds to IFN-stimulated response elements on cellular DNA, leading to the expression of the multiple ISGs. Microarray analyses show that hundreds of genes are induced by type-1 IFN, many related to antiviral activity but others involved in lipid metabolism, apoptosis, protein degradation and inflammatory cell responses²⁸.

Exogenously supplied recombinant IFN- α binds to and activates cellular receptors, leading to the same response cascades that occur with endogenous production. Hence, it has been assumed that IFN treatment works by similar mechanisms to endogenous IFN, with the greater effectiveness being caused by the higher concentrations achieved. The lack of a cell-culture system or small-animal model of HCV has made the study of therapeutic IFN difficult. Information has largely been derived from the use of subgenomic replicons and cellular protein production vector systems and, to a limited extent, from infected chimpanzees and humans.

In the replicon system, HCV RNA replication has minimal effects on the expression of ISGs, probably because of defects in IFN signalling in the cell lines that support HCV replicons^{29,30} (see the review in this issue by Gale and Foy, page 939). By contrast, exogenous administration of IFN to these cell systems, with and without active replicons, induces a broad array of ISGs and leads to a rapid reduction in virus levels to the limit of detection^{31–34}. IFN- α seems to affect the translation of viral proteins³⁵, compatible with the known actions of several ISGs, including the gene encoding protein kinase R (PKR), which blocks viral protein synthesis through inhibition of eukaryotic initiation factor 2 (eIF2). IFN- α might also decrease viral RNA stability through either 2',5'-oligoadenylate synthetase, which triggers ribonuclease-L activation, or other pathways³⁵. Although HCV repli-

Box 1 | Factors that influence therapy

Factors correlated with a sustained response to combination therapy with pegylated interferon and ribavirin in hepatitis C.

Viral factors

Genotypes 2 and 3 (versus genotype 1)
Lower viral levels
Greater quasispecies diversity
Acute versus chronic infection

Host factors

Female sex
Younger age
Less fibrosis
Lower body weight and body mass index
Non-African-American race
Absence of significant co-morbidities
(alcohol abuse, renal disease, HIV infection)

cons with decreased sensitivity to IFN- α have been identified, none has the degree of resistance that occurs in some patients with hepatitis C. Furthermore, strains of virus from non-responder patients are sensitive to IFN- α when studied in the replicon system^{33,34}.

Molecular and microarray data from chimpanzees acutely or chronically infected with HCV confirm the importance of IFN pathways and the innate immune system in the host response to infection^{36–38}. During acute HCV infection, > 300 genes are induced in the liver, many during the first few weeks of infection, well before the onset of clinical disease. Notably, neither type-1 or type-2 IFN-encoding genes are induced, although increases occur in the transcription of many ISGs, suggesting that only small amounts of IFN are necessary for ISG induction or that ISGs might be induced independently of IFN by the presence of viral double-stranded RNA or protein. In similar studies of chronically infected chimpanzees, gene induction was relatively uniform and included multiple ISGs³⁷. The effects of therapeutic IFN have yet to be fully evaluated in the chimpanzee model.

Humans with chronic hepatitis C demonstrate vigorous gene expression in response to IFN- α therapy in peripheral-blood mononuclear cells³⁹. The difficulty of obtaining liver tissue from humans at multiple time points limits studies of hepatic gene expression during therapy. Recently, Chen *et al.*⁴⁰ have reported microarray results from liver biopsy tissue taken before therapy in a cohort of patients given pegylated interferon and ribavirin. Patients who were subsequently identified as non-responders had high baseline expression of ISGs, whereas responders to therapy (and those who relapsed) more closely resembled healthy controls. These findings suggest that non-responders have an upregulated and largely ineffective IFN response, so that administration of exogenous IFN- α adds little. Alternatively, these results suggest that downstream inhibitors, either viral or host-related, are active, rendering both endogenous and exogenous IFN ineffective.

HCV inhibition of IFN actions

Abnormalities in the downstream actions of IFN activity mediated directly or indirectly by HCV have been suggested in several studies. Enomoto *et al.*⁴¹ found a correlation between variations in the amino-acid sequence of non-structural-protein 5A (NS5A) of HCV and sensitivity to IFN- α . Patients with multiple mutations in a 40-amino-acid region of NS5A, the so-called IFN-sensitivity-determining region (ISDR), were more likely to achieve SVR with therapy than those with few mutations or the consensus, wild-type sequence. However, subsequent studies failed to confirm this association, and the correlation between multiple mutations in the ISDR and lower baseline levels of virus might better explain the relationship with SVR⁴².

A role for the ISDR of NS5A in determining the response to IFN- α therapy has also been suggested by molecular studies. Gale *et al.*^{43,44} demonstrated that NS5A can bind to and inactivate PKR *in vitro*, and that this binding is dependent on the presence of the ISDR and an additional 26 carboxy-terminal amino acids of NS5A. Furthermore,

only wild-type and not mutant NS5A partly abrogates the IFN-induced PKR-dependent suppression of HCV protein translation⁴⁵. These findings provide a biological basis for correlations between the HCV NS5A sequence and lack of response to IFN- α therapy.

The ability of NS5A to induce interleukin-8 (IL-8) production provides another explanation for IFN resistance. Although IL-8 is a pro-inflammatory chemokine, it also interferes with IFN-induced antiviral responses, probably at a post-transcriptional level⁴⁶. Using a tetracycline-regulated NS5A expression system, Polyak *et al.*⁴⁷ showed that NS5A induced IL-8 mRNA and protein production, resulting in IFN inhibition and viral rescue. IL-8 upregulation has also been demonstrated using microarrays in NS5A-producing cell lines⁴⁸. Finally, patients with HCV have higher serum levels of IL-8, particularly IFN non-responders⁴⁹.

Other HCV proteins have been shown to interfere with IFN signalling. For example, HCV envelope 2 (E2) protein interacts with PKR, and a 12-amino-acid sequence of E2 has homology to the PKR-eIF2 α phosphorylation homology domain (PePHD), the site at which PKR activates eIF2 α ⁵⁰. The binding of E2 to PePHD inactivates PKR *in vitro*⁵¹. This region of E2 is well conserved within specific genotypes, and the homology with PePHD is greatest with E2 from genotype 1, possibly explaining the higher IFN resistance seen in genotype-1 infections.

Recently, Foy *et al.*⁵² showed that the protease activity of NS3/4A blocks the phosphorylation and activation of IRF-3, a key effector of the IFN antiviral cascade. IRF-3 activity was restored by treating cells with an HCV serine-protease inhibitor. In addition, NS3/4A protease activity ablates HCV-induced signalling of the IFN- β receptor by retinoic-acid-inducible gene I (RIG-I) and, furthermore, HCV protease inhibition restores RIG-I function^{32,53}. These observations suggest that the HCV protease cleaves active sites in mediators of IFN action and that combination treatment using a serine-protease inhibitor and IFN- α might have important synergistic activity.

The effects of HCV proteins on other classical antiviral pathways have also been examined. Most studies have consistently shown normal levels and activity of 2',5'-oligoadenylate synthetase and MxA in the setting of HCV protein or replicon expression⁵⁴. Both the full polyprotein and the HCV core protein have been shown to inhibit the JAK-STAT pathway in cell culture and transgenic mice, although the downstream mediators of the JAK-STAT pathway were unaffected^{55–57}. Interactions with non-classical pathways of IFN action might also occur⁵⁸. These studies highlight the complexity of teasing out the actions of HCV proteins on the often-multitiered IFN signalling pathways during acute and chronic infection, and during therapy of hepatitis C^{59,60}.

Immunomodulation

In addition to its direct antiviral actions, IFN has important interactions with the adaptive and innate immune responses. Type-1 IFNs promote memory T-cell proliferation, prevent T-cell apoptosis and stimulate natural-killer-cell activation and dendritic-cell maturation⁶¹. IFN also upregulates the production of major histocompatibility complex (MHC) class-I and class-II peptides, and might promote a T-helper-1 (T_H1) over a T-helper-2 (T_H2) phenotype. In addition to direct immune stimulation, by decreasing HCV RNA replication, IFN might prevent immune exhaustion and enhance the adaptive HCV-specific immune response⁶².

Analyses of immune responses during therapy of hepatitis C have yielded discrepant results. Kamal *et al.*⁶³ found that pegylated interferon therapy, with or without ribavirin, enhanced HCV-specific CD4⁺ T-cell responses in patients who cleared HCV compared with their own baseline, with untreated patients and with those with chronic evolution. By contrast, Rahman *et al.*⁶⁴ found that T-cell responses waned during successful therapy of acute hepatitis C and that patients with viral breakthrough maintained a strong HCV-specific response. Similarly, pegylated interferon therapy of chronic hepatitis C has been reported to be associated with enhanced T-cell responses in sustained responders in some reports, whereas others have found no induction of T-cell responses regardless of treatment outcome^{65,66}. Immune responses dur-

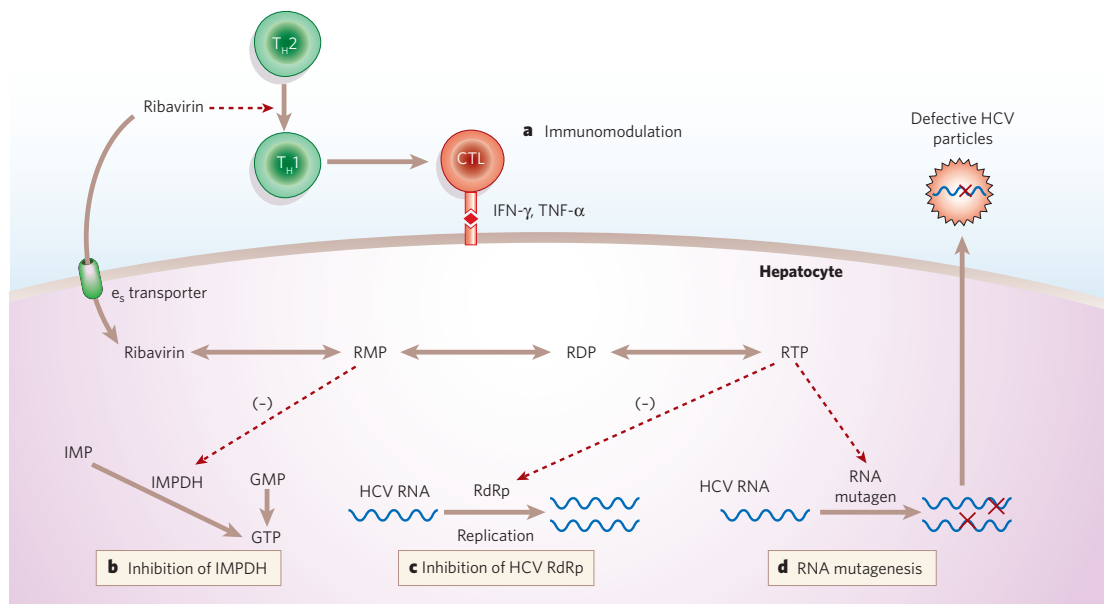


Figure 3 | Proposed mechanisms by which ribavirin could act in HCV infection. These include **a**, immunomodulation promoting T_H1 over T_H2 phenotype, **b**, IMPDH inhibition leading to GTP depletion, **c**, direct

inhibition of HCV RNA polymerase and **d**, mutagenesis resulting in reduced virion infectivity. IMPDH, inosine monophosphate dehydrogenase; T_H , T helper cell. TNE, tumour necrosis factor.

ing therapy have been measured using peripheral blood only, which might or might not reflect T-cell responses occurring in the liver⁶⁷. Compartmentalization of relevant HCV responses in the liver might partly account for the discrepant results and the lack of clear correlation between responses to IFN therapy and T-cell responses to HCV antigens. Alternatively, the immunomodulatory effects of IFN- α might be less important than its antiviral effects in treating hepatitis C.

Ribavirin

Initially synthesized as a guanosine analogue in 1970, ribavirin was immediately recognized to possess activity against several RNA and DNA viruses. Ribavirin was first approved for use in humans as a treatment for severe respiratory syncytial virus (RSV) infection in children. Its broad antiviral activity led to trials of ribavirin monotherapy for the newly discovered HCV in the early 1990s. Ribavirin monotherapy was associated with improvements in serum aminotransferase levels in at least half of patients, but viral levels did not change and patients did not clear HCV even with prolonged treatment^{68,69}. Surprisingly, the addition of ribavirin to IFN- α therapy led to marked improvements in SVR rates, increasing the proportion of patients who cleared the virus and also decreasing the relapse rate⁴. Ribavirin was subsequently approved for use in chronic hepatitis C, but only as a combination therapy with IFN- α .

Pawlotsky *et al.*⁷⁰ recently reassessed the effects of ribavirin monotherapy on early viral kinetics. Ribavirin led to a small, early, transient reduction in HCV viraemia in a proportion of patients. When used in combination, ribavirin had no effect on the first and second phases of viral kinetics but did reduce the rebound in viral levels seen before the second dose of IFN. These effects correlated with ribavirin concentration and elimination half-life. Corroborating the importance of dose, Lindahl *et al.*⁷¹ showed that, if high doses of ribavirin were used to achieve concentrations of 15 μ M, high rates of SVR (90%) could be achieved even in patients with genotype-1 infection and high viral load. Not surprisingly, toxicity was also greater. These studies illustrate the need to develop ribavirin-like agents that are better tolerated. How ribavirin augments the response rate to IFN is not known, but multiple mechanisms have been proposed, each with some experimental support (Fig. 3).

Direct inhibition of HCV replication

As a guanosine analogue, ribavirin is phosphorylated intracellularly to

form the monophosphate (RMP), diphosphate (RDP) and triphosphate (RTP). The misincorporation of RTP by RNA polymerases could lead to early chain termination and inhibition of replication. Indeed, RTP has been shown to be a weak inhibitor of many viral polymerases, including that of bovine diarrhoeal virus, a virus closely related to HCV⁷². Using an HCV RNA-dependent RNA-polymerase assay, Maag *et al.*⁷³ showed that RTP was incorporated into nascent viral RNA opposite cytosine or uridine, resulting in a significant block to RNA elongation. This inhibitory effect was present for polymerases from all six HCV genotypes but required fairly high concentrations (50–150 μ M) compared with the concentrations achieved in clinical use (10 μ M). Thus, although ribavirin might have a small direct effect on HCV-RNA replication through polymerase inhibition, this is unlikely to be its major mechanism of action against hepatitis C.

Inosine-monophosphate-dehydrogenase inhibition

Intracellularly, RMP is a competitive inhibitor of inosine monophosphate dehydrogenase (IMPDH), which leads to depletion of the GTP necessary for viral RNA synthesis. In the replicon system, ribavirin and other IMPDH inhibitors (mycophenolic acid and VX-497) partly inhibit HCV replication. The addition of excess guanosine abolishes the activity of both mycophenolic acid and VX-497 but only partly reverses the effects of ribavirin⁷². These findings are consistent with the minimal effects of ribavirin monotherapy on serum levels of HCV RNA and indicate that IMPDH inhibition and GTP depletion might contribute to, but are unlikely to be the major determinants of, the effects of ribavirin therapy in hepatitis C.

Mutagenesis and error catastrophe

HCV circulates in serum as many quasispecies (virions with minor genomic differences). Quasispecies diversity is caused by the high frequency of mutations that occur during viral replication owing to the poor fidelity and lack of proofreading activity of the HCV RNA polymerase. Crotty *et al.*^{74,75} introduced the concept that ribavirin acts as a viral mutagen, causing a higher frequency of mutations and pushing viruses toward the threshold of 'error catastrophe'. Several findings *in vitro* and *in vivo* support this explanation for the effects of ribavirin in hepatitis C.

In the replicon system, although ribavirin has little effect on levels of HCV replication, it significantly reduces the efficiency with which progeny subgenomic replicons transfect new cells^{29,76}, an indirect

reflection of a reduction in the quality (as opposed to quantity) of virus. Using a full-length expression system, Contreras *et al.*⁷⁷ showed that ribavirin increased the mutation frequency of HCV, with the highest rates of mutations being found in the NS5A-encoding region. Furthermore, Lanford *et al.*⁷⁸ showed that GBV-B virus (a close relative of HCV) grown in the presence of ribavirin had reduced specific infectivity. This effect was partly blocked by the addition of guanosine, suggesting that depletion of GTP by IMPDH inhibition augmented the mutagenic effect of ribavirin. At clinically relevant concentrations, only with GTP depletion could ribavirin be incorporated often enough to have an effect on the overall mutation rate.

The mutational ability of ribavirin has been examined in humans receiving ribavirin monotherapy. Young *et al.*⁷⁹ reported modest increases in mutations in circulating HCV RNA during prolonged ribavirin therapy, whereas two more recent studies found no increase in mutational frequency compared with a placebo or non-treatment^{70,80}. Thus, studies in humans have not confirmed the role of lethal mutagenesis in explaining the antiviral effects of ribavirin in hepatitis C, but the timing of assessment of mutations in these studies might not have been optimal for identifying increased mutagenesis.

Studies of viral kinetics comparing IFN- α alone with its combination with ribavirin provide some support for the hypothesis of lethal mutagenesis. The addition of ribavirin to IFN- α has little effect on early viral kinetics, except perhaps in patients with relative resistance to IFN⁸¹. Using kinetic models, Dixit *et al.*⁸² proposed that ribavirin exerts its effect predominantly in the second phase of viral decay, through increased mutagenesis, resulting in a lower rate of new productively infected hepatocytes.

Lethal mutagenesis is an attractive hypothesis to explain the effects of ribavirin in the therapy of hepatitis C. By decreasing replicative fitness and narrowing the genomic diversity of HCV, ribavirin might reduce the ability of HCV to escape immune and antiviral pressures, and thereby increase the effectiveness of IFN.

Immunomodulation

The balance of T_H1 and T_H2 CD4⁺ responses has proven relevant to HCV infection. Accumulating evidence has shown that an early T_H1 immune response leads to viral clearance, whereas a T_H2 response favours chronic evolution⁶⁶. Several studies have suggested that ribavirin can alter the T_H1/T_H2 balance favouring a T_H1 response and thus potentially improve treatment outcomes in hepatitis C. *In vitro* and at clinically relevant concentrations, ribavirin enhances T_H1 while inhibiting T_H2 cytokine production by stimulated T cells⁸³. In addition, patients treated with IFN- α and ribavirin have stronger HCV-specific T-cell responses than those treated with IFN- α alone and this correlates with SVR⁸⁴. Thus, ribavirin can modulate the immune system but how this is achieved and whether it is relevant to responses to treatment remain unclear.

Ribavirin and interferon signalling pathways

So far, few studies have looked at the interaction between IFN- α and ribavirin at the mechanistic level. Examining the effect of ribavirin in the RSV system using microarray techniques, Zhang *et al.*⁸⁵ found potential mechanisms for this interaction: by itself, ribavirin had little effect on gene regulation but, during RSV infection, it led to upregulation of several ISGs and enhanced STAT1 binding to DNA. Thus, in the RSV system, ribavirin seemed to increase the antiviral activity of endogenously produced IFN. In addition, others have shown that IL-8 is downregulated by ribavirin therapy⁸⁶. By extrapolation, these combined findings suggest that ribavirin might be active against HCV by its ability to augment or stabilize the intracellular mediators of IFN activity against HCV.

Future directions

Current therapies for hepatitis C are successful in 90% of patients with acute and 50% of those with chronic infection. Studies of the mechanisms of action of IFN and ribavirin and other advances in HCV

research might soon close this gap. The recent description of a cell-culture system that allows complete HCV replication will probably provide tools to characterize more fully the mechanisms of antiviral activity of IFN and the means to assess new antiviral agents^{87–89}. Furthermore, small-molecule inhibitors of the HCV protease and polymerase have been developed, and pilot studies in humans have shown them to be highly potent at lowering levels of HCV RNA^{90,91}. More importantly, perhaps, they might act synergistically by both decreasing HCV replication and interfering with the ability of HCV to evade the mediators of IFN action. These combinations hold the promise of greatly improved rates of response to therapy of hepatitis C. ■

1. Hoofnagle, J. H. *et al.* Treatment of chronic non-A, non-B hepatitis with recombinant human alpha interferon. A preliminary report. *New Engl. J. Med.* **315**, 1575–1578 (1986).
2. Lau, D. T. *et al.* 10-year follow-up after interferon-alpha therapy for chronic hepatitis C. *Hepatology* **28**, 1121–1127 (1998).
3. Di Bisceglie, A. M. & Hoofnagle, J. H. Optimal therapy of hepatitis C. *Hepatology* **36**, S121–S127 (2002).
4. McHutchison, J. G. & Poynard, T. Combination therapy with interferon plus ribavirin for the initial treatment of chronic hepatitis C. *Semin. Liver Dis.* **19** (Suppl. 1), 57–65 (1999).
5. Glue, P. *et al.* Pegylated interferon-alpha2b: pharmacokinetics, pharmacodynamics, safety, and preliminary efficacy data. Hepatitis C Intervention Therapy Group. *Clin. Pharmacol. Ther.* **68**, 556–567 (2000).
6. Zeuzem, S. *et al.* PEGinterferon α -2a in patients with chronic hepatitis C. *New Engl. J. Med.* **343**, 1666–1672 (2000).
7. Heathcote, E. J. *et al.* PEGinterferon α -2a in patients with chronic hepatitis C and cirrhosis. *New Engl. J. Med.* **343**, 1673–1680 (2000).
8. Lindsay, K. L. *et al.* A randomized, double-blind trial comparing PEGylated interferon alfa-2b to interferon α -2b as initial treatment for chronic hepatitis C. *Hepatology* **34**, 395–403 (2001).
9. Manns, M. P. *et al.* PEGinterferon α -2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* **358**, 958–965 (2001).
10. Fried, M. W. *et al.* PEGinterferon α -2a plus ribavirin for chronic hepatitis C virus infection. *New Engl. J. Med.* **347**, 975–982 (2002).
11. Hadziyannis, S. J. *et al.* PEGinterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann. Intern. Med.* **140**, 346–355 (2004).
12. Strader, D. B. Understudied populations with hepatitis C. *Hepatology* **36**, S226–S236 (2002).
13. Marcellin, P. *et al.* Long-term histologic improvement and loss of detectable intrahepatic HCV RNA in patients with chronic hepatitis C and sustained response to interferon-alpha therapy. *Ann. Intern. Med.* **127**, 875–881 (1997).
14. Shiffman, M. L. Retreatment of patients with chronic hepatitis C. *Hepatology* **36**, S128–S134 (2002).
15. Neumann, A. U. *et al.* Hepatitis C viral dynamics *in vivo* and the antiviral efficacy of interferon- α therapy. *Science* **282**, 103–107 (1998).
16. Layden, J. E. & Layden, T. J. Viral kinetics of hepatitis C: new insights and remaining limitations. *Hepatology* **35**, 967–970 (2002).
17. Lau, J. Y. *et al.* Distribution of hepatitis C virus genotypes determined by line probe assay in patients with chronic hepatitis C seen at tertiary referral centers in the United States. Hepatitis Interventional Therapy Group. *Ann. Intern. Med.* **124**, 868–876 (1996).
18. Muir, A. J., Bornstein, J. D. & Killenberg, P. G. PEGinterferon α -2b and ribavirin for the treatment of chronic hepatitis C in blacks and non-Hispanic whites. *New Engl. J. Med.* **350**, 2265–2271 (2004).
19. Jaekel, E. *et al.* Treatment of acute hepatitis C with interferon α -2b. *New Engl. J. Med.* **345**, 1452–1457 (2001).
20. Santantonio, T. *et al.* Efficacy of a 24-week course of PEG-interferon alpha-2b monotherapy in patients with acute hepatitis C after failure of spontaneous clearance. *J. Hepatol.* **42**, 329–333 (2005).
21. Bekisz, J., Schmeisser, H., Hernandez, J., Goldman, N. D. & Zoon, K. C. Human interferons alpha, beta and omega. *Growth Factors* **22**, 243–251 (2004).
22. Sen, G. C. Viruses and interferons. *Annu. Rev. Microbiol.* **55**, 255–281 (2001).
23. Robek, M. D., Boyd, B. S. & Chisari, F. V. Lambda interferon inhibits hepatitis B and C virus replication. *J. Virol.* **79**, 3851–3854 (2005).
24. Frese, M. *et al.* Interferon-gamma inhibits replication of subgenomic and genomic hepatitis C virus RNAs. *Hepatology* **35**, 694–703 (2002).
25. Soza, A. *et al.* Pilot study of interferon gamma for chronic hepatitis C. *J. Hepatol.* **43**, 67–71 (2005).
26. Gilmour, K. C. & Reich, N. C. Signal transduction and activation of gene transcription by interferons. *Gene Expr.* **5**, 1–18 (1995).
27. Gale, M. Jr Effector genes of interferon action against hepatitis C virus. *Hepatology* **37**, 975–978 (2003).
28. de Veer, M. J. *et al.* Functional classification of interferon-stimulated genes identified using microarrays. *J. Leukocyte Biol.* **69**, 912–920 (2001).
29. Lanford, R. E. *et al.* Antiviral effect and virus-host interactions in response to alpha interferon, gamma interferon, poly(i)-poly(c), tumor necrosis factor alpha, and ribavirin in hepatitis C virus subgenomic replicons. *J. Virol.* **77**, 1092–1104 (2003).
30. Sumpter, R. Jr *et al.* Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *J. Virol.* **79**, 2689–2699 (2005).
31. Blight, K. J., Kolykhalov, A. A. & Rice, C. M. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**, 1972–1974 (2000).
32. Guo, J. T., Bichko, V. V. & Seeger, C. Effect of alpha interferon on the hepatitis C virus replicon. *J. Virol.* **75**, 8516–8523 (2001).

33. Abe, K. *et al.* cDNA microarray analysis to compare HCV subgenomic replicon cells with their cured cells. *Virus Res.* **107**, 73–81 (2005).
34. Zhu, H. *et al.* Gene expression associated with interferon α antiviral activity in an HCV replicon cell line. *Hepatology* **37**, 1180–1188 (2003).
35. Guo, J. T., Sohn, J. A., Zhu, Q. & Seeger, C. Mechanism of the interferon alpha response against hepatitis C virus replicons. *Virology* **325**, 71–81 (2004).
36. Bigger, C. B. *et al.* Intrahepatic gene expression during chronic hepatitis C virus infection in chimpanzees. *J. Virol.* **78**, 13779–13792 (2004).
37. Bigger, C. B., Brasky, K. M. & Lanford, R. E. DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. *J. Virol.* **75**, 7059–7066 (2001).
38. Su, A. I. *et al.* Genomic analysis of the host response to hepatitis C virus infection. *Proc. Natl Acad. Sci. USA* **99**, 15669–15674 (2002).
39. Ji, X. *et al.* Interferon α regulated gene expression in patients initiating interferon treatment for chronic hepatitis C. *Hepatology* **37**, 610–621 (2003).
40. Chen, L. *et al.* Hepatic gene expression discriminates responders and nonresponders in treatment of chronic hepatitis C viral infection. *Gastroenterology* **128**, 1437–1444 (2005).
41. Enomoto, N. *et al.* Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *New Engl. J. Med.* **334**, 77–81 (1996).
42. Herion, D. & Hoofnagle, J. H. The interferon sensitivity determining region: all hepatitis C virus isolates are not the same. *Hepatology* **25**, 769–771 (1997).
43. Gale, M. J. Jr *et al.* Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology* **230**, 217–227 (1997).
44. Gale, M. J. Jr *et al.* Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. *Mol. Cell. Biol.* **18**, 5208–5218 (1998).
45. Wang, C. *et al.* Alpha interferon induces distinct translational control programs to suppress hepatitis C virus RNA replication. *J. Virol.* **77**, 3898–3912 (2003).
46. Khabar, K. S. *et al.* The alpha chemokine, interleukin 8, inhibits the antiviral action of interferon alpha. *J. Exp. Med.* **186**, 1077–1085 (1997).
47. Polyak, S. J. *et al.* Hepatitis C virus nonstructural 5A protein induces interleukin-8, leading to partial inhibition of the interferon-induced antiviral response. *J. Virol.* **75**, 6095–6106 (2001).
48. Girard, S. *et al.* An altered cellular response to interferon and up-regulation of interleukin-8 induced by the hepatitis C viral protein NS5A uncovered by microarray analysis. *Virology* **295**, 272–283 (2002).
49. Polyak, S. J., Khabar, K. S., Rezeiqi, M. & Gretch, D. R. Elevated levels of interleukin-8 in serum are associated with hepatitis C virus infection and resistance to interferon therapy. *J. Virol.* **75**, 6209–6211 (2001).
50. Taylor, D. R., Shi, S. T., Romano, P. R., Barber, G. N. & Lai, M. M. Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science* **285**, 107–110 (1999).
51. Pavio, N., Taylor, D. R. & Lai, M. M. Detection of a novel unglycosylated form of hepatitis C virus E2 envelope protein that is located in the cytosol and interacts with PKR. *J. Virol.* **76**, 1265–1272 (2002).
52. Foy, E. *et al.* Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. *Science* **300**, 1145–1148 (2003).
53. Foy, E. *et al.* Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-I signaling. *Proc. Natl Acad. Sci. USA* **102**, 2986–2991 (2005).
54. Frese, M., Pietschmann, T., Moradpour, D., Haller, O. & Bartenschlager, R. Interferon-alpha inhibits hepatitis C virus subgenomic RNA replication by an MxA-independent pathway. *J. Gen. Virol.* **82**, 723–733 (2001).
55. Blindenbacher, A. *et al.* Expression of hepatitis C virus proteins inhibits interferon alpha signaling in the liver of transgenic mice. *Gastroenterology* **124**, 1465–1475 (2003).
56. Basu, A., Meyer, K., Ray, R. B. & Ray, R. Hepatitis C virus core protein modulates the interferon-induced transacting factors of JAK/STAT signaling pathway but does not affect the activation of downstream IRF-1 or 561 gene. *Virology* **288**, 379–390 (2001).
57. Heim, M. H., Moradpour, D. & Blum, H. E. Expression of hepatitis C virus proteins inhibits signal transduction through the Jak-STAT pathway. *J. Virol.* **73**, 8469–8575 (1999).
58. Honda, M., Shimazaki, T. & Kaneko, S. La protein is a potent regulator of replication of hepatitis C virus in patients with chronic hepatitis C through internal ribosomal entry site-directed translation. *Gastroenterology* **128**, 449–462 (2005).
59. Katze, M. G., He, Y. & Gale, M. J. Jr. Viruses and interferon: a fight for supremacy. *Nature Rev. Immunol.* **2**, 675–687 (2002).
60. Taylor, D. R. Hepatitis C virus: evasion of the interferon-induced antiviral response. *J. Mol. Med.* **78**, 182–190 (2000).
61. Tilg, H. New insights into the mechanisms of interferon α : an immunoregulatory and anti-inflammatory cytokine. *Gastroenterology* **112**, 1017–1021 (1997).
62. Lechner, F. *et al.* Analysis of successful immune responses in persons infected with hepatitis C virus. *J. Exp. Med.* **191**, 1499–1512 (2000).
63. Kamal, S. M. *et al.* PEGylated interferon alpha therapy in acute hepatitis C: relation to hepatitis C virus-specific T cell response kinetics. *Hepatology* **39**, 1721–1731 (2004).
64. Rahman, F. *et al.* Effects of antiviral therapy on the cellular immune response in acute hepatitis C. *Hepatology* **40**, 87–97 (2004).
65. Kamal, S. M., Fehr, J., Roesler, B., Peters, T. & Rasenack, J. W. PEGinterferon alone or with ribavirin enhances HCV-specific CD4 T-helper 1 responses in patients with chronic hepatitis C. *Gastroenterology* **123**, 1070–1083 (2002).
66. Rehmann, B. & Nascimbeni, M. Immunology of hepatitis B virus and hepatitis C virus infection. *Nature Rev. Immunol.* **5**, 215–229 (2005).
67. Thomson, M. *et al.* The clearance of hepatitis C virus infection in chimpanzees may not necessarily correlate with the appearance of acquired immunity. *J. Virol.* **77**, 862–870 (2003).
68. Di Bisceglie, A. M. *et al.* Ribavirin as therapy for chronic hepatitis C. A randomized, double-blind, placebo-controlled trial. *Ann. Intern. Med.* **123**, 897–903 (1995).
69. Hoofnagle, J. H., Lau, D., Conjeevaram, H., Kleiner, D. & Di Bisceglie, A. M. Prolonged therapy of chronic hepatitis C with ribavirin. *J. Viral Hepat.* **3**, 247–252 (1996).
70. Pawlotsky, J. M. *et al.* Antiviral action of ribavirin in chronic hepatitis C. *Gastroenterology* **126**, 703–714 (2004).
71. Lindahl, K., Stahl, L., Bruchfeld, A. & Schvarcz, R. High-dose ribavirin in combination with standard dose PEGinterferon for treatment of patients with chronic hepatitis C. *Hepatology* **41**, 275–279 (2005).
72. Lau, J. Y., Tam, R. C., Liang, T. J. & Hong, Z. Mechanism of action of ribavirin in the combination treatment of chronic HCV infection. *Hepatology* **35**, 1002–1009 (2002).
73. Maag, D., Castro, C., Hong, Z. & Cameron, C. E. Hepatitis C virus RNA-dependent RNA polymerase (NS5B) as a mediator of the antiviral activity of ribavirin. *J. Biol. Chem.* **276**, 46094–46098 (2001).
74. Crotty, S., Cameron, C. E. & Andino, R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc. Natl Acad. Sci. USA* **98**, 6895–6900 (2001).
75. Crotty, S. *et al.* The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nature Med.* **6**, 1375–1379 (2000).
76. Zhou, S., Liu, R., Baroudy, B. M., Malcolm, B. A. & Reyes, G. R. The effect of ribavirin and IMPDH inhibitors on hepatitis C virus subgenomic replicon RNA. *Virology* **310**, 333–342 (2003).
77. Contreras, A. M. *et al.* Viral RNA mutations are region specific and increased by ribavirin in a full-length hepatitis C virus replication system. *J. Virol.* **76**, 8505–8517 (2002).
78. Lanford, R. E. *et al.* Ribavirin induces error-prone replication of GB virus B in primary tamarin hepatocytes. *J. Virol.* **75**, 8074–8081 (2001).
79. Young, K. C. *et al.* Identification of a ribavirin-resistant NS5B mutation of hepatitis C virus during ribavirin monotherapy. *Hepatology* **38**, 869–878 (2003).
80. Lutchman, G. A. *et al.* Mutation rate of hepatitis C virus in patients during ribavirin monotherapy. *Hepatology* **41**, 385A (2004).
81. Layden-Almer, J. E., Ribeiro, R. M., Wiley, T., Perelson, A. S. & Layden, T. J. Viral dynamics and response differences in HCV-infected African American and white patients treated with IFN and ribavirin. *Hepatology* **37**, 1343–1350 (2003).
82. Dixit, N. M., Layden-Almer, J. E., Layden, T. J. & Perelson, A. S. Modelling how ribavirin improves interferon response rates in hepatitis C virus infection. *Nature* **432**, 922–924 (2004).
83. Tam, R. C. *et al.* Ribavirin polarizes human T cell responses towards a type 1 cytokine profile. *J. Hepatol.* **30**, 376–382 (1999).
84. Cramp, M. E. *et al.* Hepatitis C virus-specific T-cell reactivity during interferon and ribavirin treatment in chronic hepatitis C. *Gastroenterology* **118**, 346–355 (2000).
85. Zhang, Y. *et al.* Ribavirin treatment up-regulates antiviral gene expression via the interferon-stimulated response element in respiratory syncytial virus-infected epithelial cells. *J. Virol.* **77**, 5933–5947 (2003).
86. Fiedler, M. A., Wernke-Dollries, K. & Stark, J. M. Inhibition of viral replication reverses respiratory syncytial virus-induced NF- κ B activation and interleukin-8 gene expression in A549 cells. *J. Virol.* **70**, 9079–9082 (1996).
87. Wakita, T. *et al.* Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Med.* **11**, 791–796 (2005).
88. Zhong, J. *et al.* Robust hepatitis C virus infection *in vitro*. *Proc. Natl Acad. Sci. USA* **102**, 9294–9299 (2005).
89. Lindenbach, B. D. *et al.* Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623–626 (2005).
90. Reesink, H. W. *et al.* Initial results of a phase 1B, multiple-dose study of VX-950, a hepatitis C virus protease inhibitor. *Gastroenterology* **128**, 527A (2005).
91. Rodriguez-Torres, M. *et al.* Valopicitabine (nm283) alone and in combination with PEG-interferon in patients with genotype 1 chronic hepatitis C: preliminary results from an ongoing phase II, multicenter study. *Gastroenterology* **128**, 695A (2005).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H.H. (Hoofnagle@extra.niddk.nih.gov).

Hepatitis C and liver transplantation

Robert S. Brown Jr¹

Liver transplantation is a life-saving therapy to correct liver failure, portal hypertension and hepatocellular carcinoma arising from hepatitis C infection. But despite the successful use of living donors and improvements in immunosuppression and antiviral therapy, organ demand continues to outstrip supply and recurrent hepatitis C with accelerated progression to cirrhosis of the graft is a frequent cause of graft loss and the need for retransplantation. Appropriate selection of candidates and timing of transplantation, coupled with better pre- and post-transplant antiviral therapy, are needed to improve outcomes.

When antiviral therapy fails in hepatitis C virus (HCV) infection, or if diagnosis of the disease is delayed until the appearance of decompensated liver disease with portal hypertension, the only option for the individual is liver replacement. Currently, over 17,000 individuals are awaiting orthotopic liver transplantation, and fewer than 5,000 liver transplants are performed per year (for the latest data see the Organ Procurement and Transplantation Network website at <http://www.optn.org/data>). This represents the greatest challenge in liver transplantation: namely, the demand for organs vastly outstrips the supply.

Hepatitis C is the most common indication for orthotopic liver transplantation, accounting for 40–50% of both individuals on the waiting list and those who have undergone liver transplants. Thus, there are insufficient donor organs even if only transplant candidates with HCV are considered. Unfortunately, liver transplantation is not a cure for hepatitis C. Viral recurrence is universal and damage to the new liver occurs routinely. Recurrent HCV infection is among the leading causes of graft loss and the need for retransplantation. Thus, the challenges in liver transplantation as a treatment for hepatitis C include accessing adequate numbers of liver grafts and controlling the virus before and after transplantation to mitigate recurrent disease.

Listing for transplantation

After evaluation, acceptable candidates for liver transplantation in the United States are registered with the United Network for Organ Sharing (UNOS). This organization runs a centralized computer network that includes the waiting list of every transplant hospital and that links all organ procurement organizations. Organs are allocated first locally in an organ procurement organization, and then regionally in the 11 UNOS areas, and finally nationwide for individuals with chronic liver disease.

Prioritization on the waiting list

Organ allocation was previously based principally on location (whether the individual was at home, in hospital or in intensive care) and on waiting time. Allocation has recently shifted to a risk-based priority system that uses MELD, a mathematical model for end-stage liver disease. MELD is based on logarithmic transformation of the potential recipient's INR (a measure of blood clotting), bilirubin and creatinine. MELD predicts short-term mortality for those on the waiting list more accurately than does the Child–Pugh score (a scoring system for severity of cirrhosis), which was previously used in the organ allocation scheme.

Within a distribution unit, individuals with the highest MELD

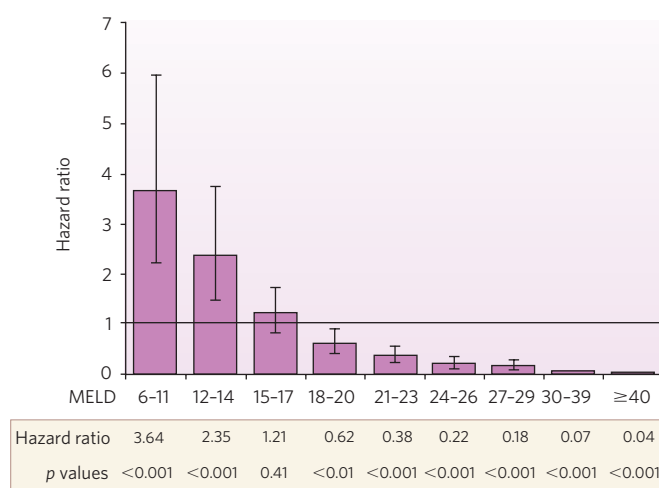


Figure 1 | The mathematical model MELD can predict the survival of candidates waiting for a liver transplant. Individuals who have MELD scores higher than 15 derive a progressively increasing survival benefit with each subsequent increment in their score⁴. The hazard ratio is defined as the likelihood of death during the year after undergoing transplantation compared with remaining on the waiting list. A hazard ratio of more than 1 means that a patient is more likely to die with transplantation and a ratio of less than 1 means the patient is less likely to die with transplantation.

score (which ranges from 6 to 40) have the highest priority for transplantation, and waiting time is used only to discriminate between individuals with the same MELD score. Currently, however, the MELD score of transplant recipients varies widely in the different organ procurement organizations. Donor organs should be allocated to individuals who are most likely to benefit from a transplant¹. In the first 18 months after the MELD-based allocation system was introduced, the overall pre-transplant mortality decreased². Recipient and graft survival also increased after the MELD model was implemented.

Individuals with hepatitis C are allocated organs according to their MELD scores, but they receive additional priority if they develop hepatocellular carcinoma (HCC). Additional MELD points are given to all individuals with HCC because laboratory MELD scores do not reflect the mortality risk from this disease. In the previous allocation scheme, the low priority given to these individuals led to a high

¹Departments of Medicine and Surgery, and Center for Liver Disease and Transplantation, Columbia University College of Physicians and Surgeons, 622 West 168th Street, New York, New York 10032, USA.

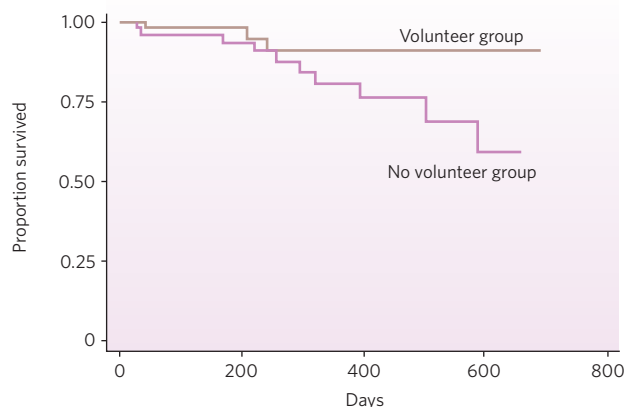


Figure 2 | Survival benefit of pursuing live donor liver transplantation (LDLT). Monitoring the survival of candidates on the transplant waiting list with (volunteer group) and without (no volunteer group) a potential donor shows the benefit of LDLT¹⁰.

dropout rate caused by tumour progression. Because individuals with HCC benefit from early transplantation before metastasis or growth of the lesion, those with small tumours (fewer than three lesions < 3 cm or one lesion < 5 cm) are given added priority: they start with a minimum MELD score of 22 and their score is increased every 3 months. No priority is given to those with tumours that exceed these size limits owing to the increased risk of recurrent HCC after transplantation. This change in allocation has resulted in a marked increase in the proportion of individuals with HCC receiving a transplant, and has decreased both the waiting time and the dropout rate. The availability of accelerated transplantation has increased the value of screening for HCC in individuals with HCV-related cirrhosis.

Timing of transplantation

Owing to regional variances in the criteria for placing an individual on the waiting list for liver transplantation, in 1997 the American Society of Transplant Physicians and the American Association for the Study of Liver Diseases published recommendations regarding the minimum criteria that an adult should meet to be put on this list³. These guidelines recommend that an individual placed on the waiting list should be ready to proceed with transplantation immediately should an organ become available. To qualify for listing, the individual's expected chance of surviving 1 year without transplantation should be 90% or less. On the basis of published data, individuals with a Child–Pugh score of seven or more, or those with bleeding associated with portal hypertension, meet these criteria and should be evaluated and listed for transplantation.

The timing of transplantation involves determining when an individual will derive the maximum benefit from receiving a new liver. The goal is to avoid both premature transplants when liver disease is not advanced and futile transplants when individuals are too sick. On the one hand, if the transplant is performed before liver failure develops, then the morbidity and mortality of the transplant operation will outweigh the benefits. This is particularly true for hepatitis C, for which delaying the initiation of recurrent HCV in the new graft may add years to an individual's life and may allow time for the development of a new antiviral therapy. Thus, early transplantation is potentially more harmful for these individuals, unless it is linked to pre-transplant antiviral therapy and HCV eradication (see below). On the other hand, if a transplant candidate is moribund, then the surgical risks of the procedure can become prohibitive.

A recent study⁴ suggests that those with MELD scores of less than 15, particularly less than 12, do not derive a survival benefit in the first year, whereas the survival benefit increases with each increment in score for those with MELD scores of more than 15 (Fig. 1). At high MELD scores (>30), the risk of dying after transplantation was found to increase by 50% and more individuals were removed from the wait-

ing list for the reason of 'death' or 'too sick', but outcomes in the sickest individuals were still reasonable, especially given the over 300-fold increase in pre-transplant mortality in candidates with high MELD scores⁴. Other studies indicate that the MELD score is a relatively poor predictor of post-transplant outcomes in all but individuals with the highest 20% of MELD scores⁵.

However, retransplant candidates, individuals with renal failure requiring dialysis and those requiring mechanical ventilation, particularly older individuals, have a significantly increased risk of operative mortality⁵. The presence of two or more risk factors predicts a very low post-transplant survival. Similarly, in living donor transplant recipients, analysis of UNOS data has shown that being in the intensive care unit before transplant, retransplantation, female donor to male recipient transplantation, being 44 yr or older and of non-white recipient race can increase the rate of retransplantation, but not death⁶. Thus, there is no absolute cut-off in MELD score for transplantation futility.

Thus, the optimal time for liver transplantation is when an individual achieves a MELD score of 15 or more or begins to show evidence of decompensation, manifested by synthetic dysfunction, or malnutrition. Although prioritization for orthotopic liver transplantation is not affected by an early referral, it does allow pre-transplant problems to be addressed and the management and timing of transplantation to be optimized.

If an individual would not derive a survival benefit from transplantation—either because their condition has worsened such that the procedure's risks outweigh its benefits or, rarely, because the individual's condition has improved—it is appropriate to remove them from the list permanently or temporarily.

Living donor liver transplantation and hepatitis C

First performed in children in 1989 (ref. 7), living donor liver transplantation (LDLT) has been performed from adult to adult in the United States since 1998 (ref. 8). Because living donation permits transplantation to take place independent of either waiting time or the severity of liver disease, the criteria required for LDLT differ from those required for deceased donor liver transplantation (DDLT). Because a living donor organ has significantly less cold ischaemia time than does a deceased donor organ because it is transferred immediately from donor to recipient, and because it is from a healthy, extensively screened individual, living donor livers are potentially of better quality than are deceased donor livers. The living donor allograft, however, has significantly less hepatic mass than has a full-sized deceased donor organ. So far, the outcomes of living donation and deceased donor transplantation have been similar.

The reduced waiting period for a living donor organ — the principal benefit of living donor transplants — may decrease the risks of decompensation or death before transplantation, thereby improving the overall chances for success. Data show that individuals on the waiting list with potential donors for LDLT have improved survival: their mortality is half that of those listed only for DDLT^{9,10} (Fig. 2). Because the transplant is done on an elective basis, the operation can proceed immediately after the workup. Alternatively, the flexibility of the waiting period before transplantation in living donor recipients can allow an attempt at pre-transplant viral eradication. It seems that if the recipient is negative for serum HCV RNA on therapy, then LDLT leads to a very low percentage (10%) of post-transplant viral recurrence¹¹. This can facilitate a cure for hepatitis C through transplantation — an important issue because over half the individuals on the waiting list who have been previously treated cannot tolerate a full course of therapy or relapse (see below). Thus, viral eradication may be an indication for earlier transplantation by LDLT in individuals with HCV at a stage when they can tolerate antiviral therapy.

LDLT grafts have tremendous growth potential: the graft generates over 150,000 hepatocytes every second in the first week after transplantation and doubles in size within 4 weeks^{12,13}. Concerns have been raised about the effect of accelerated growth that follows LDLT grafts. Theoretically, this growth potential may predispose individuals trans-

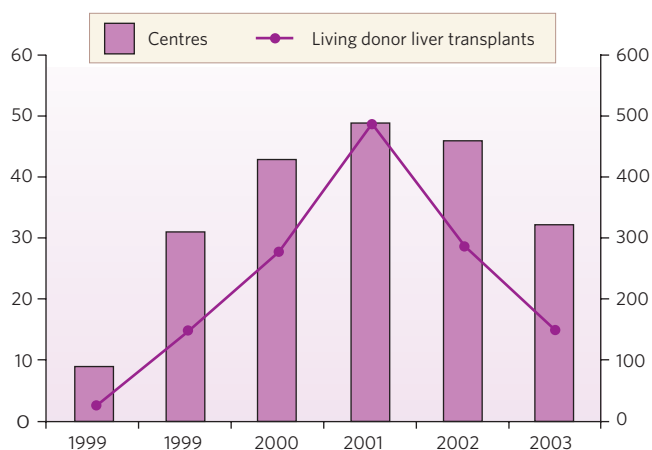


Figure 3 | Number of live donor liver transplants in the United States. Shown are the number of centres performing LDLT and the number of adult LDLT operations performed from 1998 to 2003 in the United States²¹.

planted for HCV cirrhosis to a more aggressive recurrence of HCV. Two large studies have shown that the incidence and severity of HCV recurrence do not differ between DDLT and LDLT recipients^{14,15}; however, another study has found that the incidence of cholestatic hepatitis, a particularly virulent and rapidly destructive form of recurrent HCV, is significantly greater in LDLT recipients¹⁶. Recently, a careful comparison of protocol liver biopsies from 23 LDLT and 53 DDLT recipients found no significant differences in the degree of hepatic inflammation between the two groups over 3 years, and similar or less fibrosis in the LDLT group, which reached a plateau after 12 months¹⁵.

Individuals with decompensated cirrhosis who meet the standard indications for orthotopic liver transplantation do not have any contraindications and have MELD scores of 15 or higher are the most appropriate candidates for LDLT. A simple rule is that an appropriate LDLT candidate is one who would undergo transplantation immediately if organs were unlimited. MELD can help to identify individuals who are not likely to benefit from LDLT because they are either too sick or too well to undergo transplantation¹⁷; however, the number of potential LDLT operations is limited. In one centre, 51 out of 100 individuals evaluated for LDLT were rejected¹⁸. The most frequent reasons for rejection included medical co-morbidity, high-risk psychosocial issues, obesity, financial issues and the procurement of a deceased donor organ during the evaluation¹⁸. Overall, in experienced centres about a third of adults on the waiting list may have a potential donor and half of these will undergo the procedure; thus, LDLT may be applicable in up to 15% of individuals on the list¹⁹. Between 2001 and 2003, however, the number of centres performing the procedure and the number of LDLT cases dropped markedly^{8,20,21}, and currently less than 5% of all adult liver transplants use living donors (Fig. 3). This reluctance to perform LDLT may be related to two highly publicized donor deaths^{20,21}. With increased experience and the lessons learned from A2ALL, a living donor cohort study funded by the NIH, it is hoped that living donation will expand to meet organ demand better in the future.

Predictors of hepatitis C after liver transplantation

Recurrent HCV remains a persistent problem and a leading cause of graft loss. In individuals who have active HCV replication before transplantation, the reacquisition of viraemia after transplantation is universal. Attempts to prevent reinfection with immune globulin or other agents have not been successful²². Reinfection occurs during reperfusion of the liver allograft, and viral titres reach pre-transplant values at about 72 h (ref. 23). At steady state, the HCV viral load is, in general, ten times higher after transplantation than before. Histological recurrence with allograft hepatitis owing to HCV occurs in up to 90% of individuals by the fifth year after transplantation²⁴.

Although histological injury in the allograft owing to HCV is exceed-

ingly common, progression of hepatitis C is variable: some individuals experience indolent disease, whereas others progress rapidly to cirrhosis and liver failure. In those that develop recurrent cirrhosis after transplantation, rapid decompensation is common. Up to 42% of individuals with HCV-related cirrhosis after transplantation have been reported to develop decompensation, manifested as ascites, encephalopathy or hepatic hydrothorax, and less than 50% of individuals survive for 1 year after they develop decompensation²⁵. Thus, data indicate that the progression of hepatitis C is accelerated after orthotopic liver transplantation as compared with non-transplanted individuals.

Several factors have been associated with the increased severity of recurrent HCV and the decreased recipient survival. For example, although grafts from donors over 60 years (up to 80 years) function without a negative impact on recipient outcomes^{26,27} in individuals without HCV, the use of these grafts in HCV-positive recipients requires caution. Data suggest that there may be a more severe recurrence of HCV and a more rapid progression to cirrhosis when older donors are used²⁸. No adverse outcome has been found, however, when selected HCV-positive grafts with no significant liver disease are used^{29–31} or when grafts positive for hepatitis B core antibody but negative for surface antigen are used³⁰.

Of recipient factors, higher HCV viral loads before transplantation correlate with lower recipient survival after transplantation. For example, individuals with an HCV RNA titre of more than 1×10^6 copies per ml before transplantation were found to have a cumulative 5-year survival of 57% as compared with 84% for those with HCV RNA titres of less than 1×10^6 copies per ml (ref. 32). It is not known, however, whether reducing viral load will improve these outcomes. Research is currently focused on developing antiviral strategies to reduce or to eliminate the pre-transplant viral burden to lessen post-transplant recurrence. In addition, advanced recipient age, hyperbilirubinaemia, increased INR and pre-transplant cytomegalovirus status adversely affect survival after transplantation³³. Whether factors such as obesity or alcohol use accelerate histological progression after transplantation has not been well studied, but their effects are likely to be similar to those in the non-transplant setting.

Immunosuppression and HCV recurrence

Because hepatitis C progresses more rapidly after transplantation, the choice and extent of immunosuppression have been an area of active research and controversy. Standard post-transplant immunosuppression consists of a calcineurin inhibitor (tacrolimus or cyclosporine) and a tapering dose of corticosteroids with or without an anti-proliferative agent for lymphocytes (mycophenolate mofetil or azathioprine). Less frequently, antibodies to T cells or to the interleukin-2 receptor are used initially as part of an induction protocol. Data supporting the superiority of any given baseline immunosuppressive agent are limited. Most studies have shown that the severity of recurrent HCV is similar whether cyclosporine- or tacrolimus-based immunosuppression is used^{32,34–36}. Data on mycophenolate mofetil are conflicting: some reports show improved outcomes, whereas others show worse outcomes^{37,38}. Most results indicate that it is the overall intensity of immunosuppression that affects outcomes: more intense immunosuppression leads to worse outcomes. Thus, HCV-induced graft failure, progression to cirrhosis and severe cholestatic hepatitis are more common in recipients who receive high-dose bolus steroids and anti-lymphocyte and anti-interleukin-2 receptor antibody preparations^{32,39}.

These agents, however, are usually used to treat organ rejection. Treatment for rejection has been associated with diminished survival in HCV-positive but not HCV-negative recipients⁴⁰. Because almost all individuals have some degree of recurrent HCV with portal inflammation, differentiating between rejection in the setting of recurrent HCV and HCV recurrence alone can be difficult on biopsy and requires a skilled hepatopathologist. Certain features (such as lobular activity and interface hepatitis) are more compatible with hepatitis C, whereas others (such as bile duct damage and mixed cellular infil-

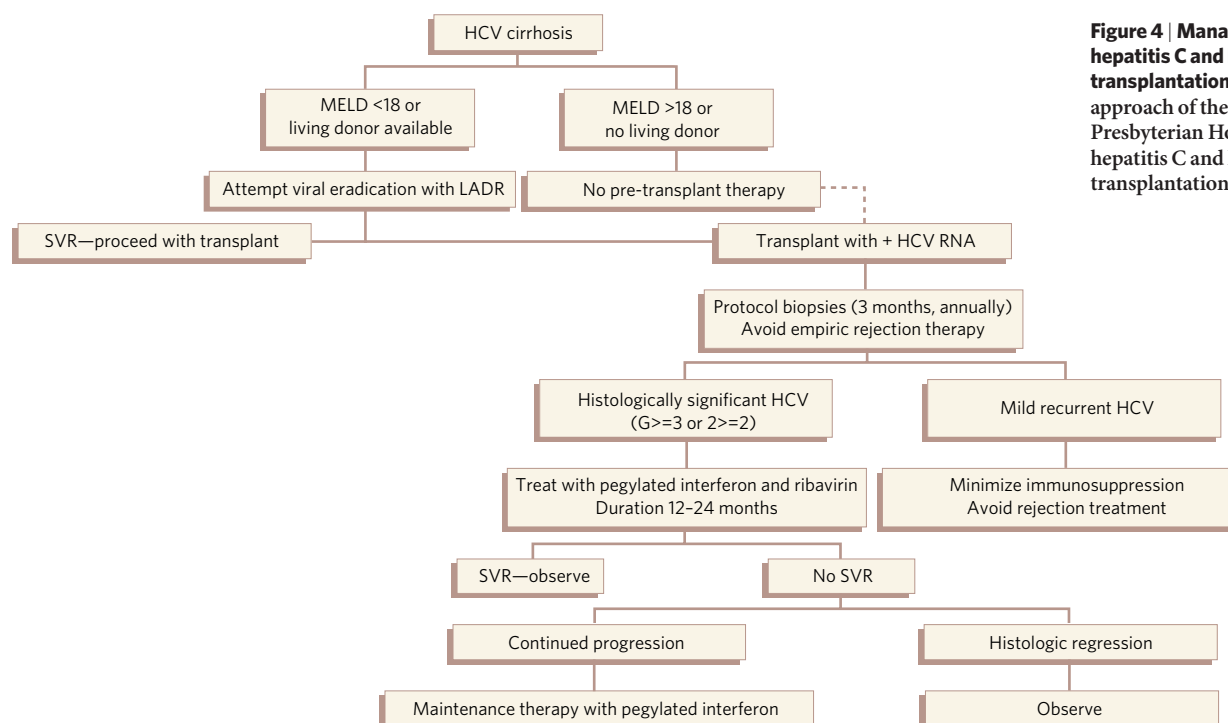


Figure 4 | Management of hepatitis C and liver transplantation. Shown is the approach of the New York Presbyterian Hospital to hepatitis C and liver transplantation.

trates) suggest rejection, but there is considerable overlap. Therefore, modulating immunosuppression in the setting of suspected mild rejection by either increasing the dose or substituting the calcineurin inhibitor and/or reintroducing mycophenolic acid may be preferable to using bolus glucocorticoids and/or antibodies to T cells.

Rapidly tapering doses of steroids and steroid-free immunosuppression with or without induction antibodies have been thought to reduce the likelihood of severe recurrent HCV. The latter may be preferable because high-dose steroids are avoided completely and preliminary data support its use⁴¹. The emerging consensus is that rapid changes in immunosuppressive level are most deleterious because they facilitate increased viral replication during intense immunosuppression, followed by immune recognition and clearance of virally infected allograft cells during rapid immunosuppressive withdrawal. Thus, our approach at New York Presbyterian Hospital has been to choose adequate immunosuppression to minimize the incidence of rejection, followed by a very gradual taper, and to avoid intense treatment for rejection with bolus steroids or antibodies. We use a calcineurin inhibitor (either cyclosporine or tacrolimus) and a slow taper of steroids over 6–12 months with mycophenolate mofetil for the first year, and take protocol biopsies at months 3, 12 and 24 to guide our decisions on immunosuppressive and antiviral treatment (Fig. 4).

Retransplantation for recurrent HCV

Because the recurrence of HCV is often accelerated after transplantation, the issue of whether to retransplant individuals with graft failure caused by recurrent HCV is highly controversial. The approach to retransplantation for recurrent HCV varies widely, and some centres no longer perform the procedure owing to poor recipient and graft survival. Individuals undergoing retransplantation for HCV have worse outcomes than do those undergoing primary transplantations; however, the outcomes are not clearly worse than those after retransplantation for other causes. Thus, it does not seem reasonable to exclude all individuals with recurrent HCV from retransplantation. Those with early, aggressive recurrence and graft failure within the first year, however, have very poor outcomes after retransplantation, as do those with very high MELD scores. These individuals should not undergo repeat transplantation except under highly selected conditions.

Treatment of hepatitis C in the peri-transplant period

Both the optimal timing and method of treating recurrent HCV after liver transplantation have been studied inadequately, but treating individuals when they are on the waiting list and pre-transplant viral eradication, respectively, represent the ideal. With pegylated interferon and ribavirin therapy, individuals with compensated cirrhosis were found to have an end-of-treatment viral response and a sustained viral response (SVR) of 23% and 11%, respectively, in the NIH-sponsored HALT-C trial⁴².

By contrast, the treatment of individuals with decompensated liver disease, who comprise most potential transplant recipients, has been far less promising. This strategy has been associated with exacerbation of encephalopathy, infection and other serious, adverse events with up to 10% mortality, as well as a low SVR⁴³. However, an initial therapy of low-dose interferon (including pegylated interferon preparations) and ribavirin, followed by a slow escalation in dose, may be associated with improved tolerability and efficacy in individuals with compensated cirrhosis⁴². This strategy has been associated with a lower incidence of adverse events, but with a discontinuation rate of 27% (ref. 44). The on-treatment viral response and SVR were reported to be 39% and 20%, respectively, in 91 subjects⁴⁴. A preliminary study of ten individuals suggests that those who achieve an SVR, or who are transplanted while on therapy with an undetectable viral load, have a less than 10% likelihood of HCV recurrence⁴⁵. This is particularly useful with LDLT because it allows timed transplantation during therapy in individuals with lower MELD scores after viral clearance in the serum without a risk of post-treatment relapse. This approach could potentially cure about 40% of individuals with HCV who undergo LDLT.

After liver transplantation, both pre-emptive therapy before the development of histological injury and directed therapy after injury occurs have been attempted with varying success. After transplantation, the tolerability of interferon preparations and ribavirin is suboptimal: significant leukopenia and anaemia are common and multifactorial, both arising from drug-induced bone marrow suppression and renal insufficiency, which potentiates ribavirin-induced haemolysis⁴⁶. Pre-emptive strategies using standard interferon and ribavirin have been associated with an on-therapy viral clearance of 23–40%, a sustained viral clearance of about 20% and a discontinuation rate of 12–50% (refs 47–49). A recent randomized trial of

pegylated interferon and ribavirin showed an SVR of only 8% with an early discontinuation rate of 31% (ref. 50). Treatment of established recurrent HCV has yielded an on-therapy response of 15–48% and an SVR of 7–26%, and has a discontinuation rate of 30–50% (refs 50–55). Longer-term treatment may improve the SVR, and many groups including ours use 18–24 months of therapy in individuals who achieve initial viral clearance.

Because sustained viral clearance is achieved in less than 30% of individuals, modulating the severity of disease and preventing graft loss are the goal. This has led to the use of maintenance therapy in many individuals, although controlled data supporting histological benefit in the absence of viral clearance are lacking. In addition, emerging data indicate that rejection increases with interferon treatment, particularly treatment associated with viral clearance^{53,55}. For example, two studies have shown acute cellular rejection in 8 out of 23 and 5 out of 44 individuals treated with interferon or pegylated interferon, coupled with a high proportion of graft cirrhosis or failure^{51,53}. In one study⁵³, four out of five individuals with rejection had viral clearance, an observation that others have noted anecdotally. This has led most groups to abandon pre-emptive therapy in favour of treating histologically significant disease.

The International Liver Transplantation Society has recommended that therapy should be initiated for all individuals with stage II fibrosis⁵⁶. Our group initiates therapy for those with grade III–IV (moderate to severe) inflammation or stage II hepatitis (Fig. 4). The use of histological triggers for initiating therapy requires careful surveillance and protocol biopsies. Research is currently focused on defining the appropriate timing, dosing and duration of treatment, and the outcomes.

Future directions

Transplantation for end-stage liver disease is a life-saving therapy to reverse the manifestations of liver failure, portal hypertension and hepatocellular carcinoma. Its application is limited primarily by the shortage of organs. Despite the successful use of grafts from older, HCV-positive and living donors, organ demand will continue to outstrip supply until adequate xenografts or hepatic stem cells can be used.

Recurrent HCV does not affect short-term survival, but the more rapid progression of disease can lead to graft loss and can lower recipient survival 3–5 years after transplantation. Increased understanding of the interaction between hepatitis C and rejection, coupled with improvements in immunosuppressive strategies, pre- and post-transplant antiviral treatments and anti-fibrotic therapy, is desperately needed to improve outcomes. Appropriate donor and graft selection, careful monitoring with protocol liver biopsies and avoidance of excess immunosuppression are crucial. Given the current suboptimal results with antiviral treatment, prevention of significant histological recurrent HCV is the preferred strategy. ■

1. Brown, R. S. Jr & Lake, J. R. The survival impact of liver transplantation in the MELD era, and the future for organ allocation and distribution. *Am. J. Transplant.* **5**, 203–204 (2005).
2. Olthoff, K. M. et al. Summary report of a national conference: evolving concepts in liver allocation in the MELD and PELD era. *Liver Transplant.* **10**, A6–A22 (2004).
3. Lucey, M. R. et al. Minimal criteria for placement of adults on the liver transplant waiting list: a report of a national conference organized by the American Society of Transplant Physicians and the American Association for the Study of Liver Diseases. *Liver Transplant. Surg.* **3**, 628–637 (1997).
4. Merion, R. M. et al. The survival benefit of liver transplantation. *Am. J. Transplant.* **5**, 307–313 (2005).
5. Desai, N. M. et al. Predicting outcome after liver transplantation: utility of the model for end-stage liver disease and a newly derived discrimination function. *Transplantation* **77**, 99–106 (2004).
6. Abt, P. L. et al. Allograft survival following adult-to-adult living donor liver transplantation. *Am. J. Transplant.* **4**, 1302–1307 (2004).
7. Broelsch, C. E. et al. Liver transplantation in children from living related donors. Surgical techniques and results. *Ann. Surg.* **214**, 428–437 (1991).
8. Brown, R. S. Jr et al. A survey of liver transplantation from living adult donors in the United States. *N. Engl. J. Med.* **348**, 818–825 (2003).
9. Liu, C. L., Lam, B., Lo, C. M. & Fan, S. T. Impact of right-lobe live donor liver transplantation on patients waiting for liver transplantation. *Liver Transplant.* **9**, 863–869 (2003).
10. Russo, M. W., LaPointe-Rudow, D., Kinkhabwala, M., Emond, J. & Brown, R. S. Jr Impact of adult living donor liver transplantation on waiting time survival in candidates listed for liver transplantation. *Am. J. Transplant.* **4**, 427–431 (2004).
11. Everson, G. T. Treatment of patients with hepatitis C virus on the waiting list. *Liver Transplant.* **9**, S90–S94 (2003).
12. Marcos, A. et al. Liver regeneration and function in donor and recipient after right lobe adult to adult living donor liver transplantation. *Transplantation* **69**, 1375–1379 (2000).
13. Baltz, A. C. & Trotter, J. F. Living donor liver transplantation and hepatitis C. *Clin. Liver Dis.* **7**, 651–665, viii (2003).
14. Russo, M. W., Galanko, J., Beavers, K., Fried, M. W. & Shrestha, R. Patient and graft survival in hepatitis C recipients after adult living donor liver transplantation in the United States. *Liver Transplant.* **10**, 340–346 (2004).
15. Shiffman, M. L. et al. Histologic recurrence of chronic hepatitis C virus in patients after living donor and deceased donor liver transplantation. *Liver Transplant.* **10**, 1248–1255 (2004).
16. Gaglio, P. J. et al. Increased risk of cholestatic hepatitis C in recipients of grafts from living versus cadaveric liver donors. *Liver Transplant.* **9**, 1028–1035 (2003).
17. Freeman, R. B. The impact of the model for end-stage liver disease on recipient selection for adult living liver donation. *Liver Transplant.* **9**, S54–S59 (2003).
18. Trotter, J. F. et al. Evaluation of 100 patients for living donor liver transplantation. *Liver Transplant.* **6**, 290–295 (2000).
19. Rudow, D. L., Russo, M. W., Hafliger, S., Emond, J. C. & Brown, R. S. Jr Clinical and ethnic differences in candidates listed for liver transplantation with and without potential living donors. *Liver Transplant.* **9**, 254–259 (2003).
20. Miller, C. et al. Fulminant and fatal gas gangrene of the stomach in a healthy live liver donor. *Liver Transplant.* **10**, 1315–1319 (2004).
21. Russo, M. W. & Brown, R. S. Jr Adult living donor liver transplantation. *Am. J. Transplant.* **4**, 458–465 (2004).
22. Charlton, M. Natural history of hepatitis C and outcomes following liver transplantation. *Clin. Liver Dis.* **7**, 585–602 (2003).
23. Garcia-Retortillo, M. et al. Hepatitis C virus kinetics during and immediately after liver transplantation. *Hepatology* **35**, 680–687 (2002).
24. Berenguer, M. Natural history of recurrent hepatitis C. *Liver Transplant.* **8**, S14–S18 (2002).
25. Berenguer, M. et al. Natural history of clinically compensated hepatitis C virus-related graft cirrhosis after liver transplantation. *Hepatology* **32**, 852–858 (2000).
26. Cescon, M. et al. Long-term survival of recipients of liver grafts from donors older than 80 years: is it achievable? *Liver Transplant.* **9**, 1174–1180 (2003).
27. Zhao, Y., Lo, C. M., Liu, C. L. & Fan, S. T. Use of elderly donors (> 60 years) for liver transplantation. *Asian J. Surg.* **27**, 114–119 (2004).
28. Machicao, V. I. et al. Donor age affects fibrosis progression and graft survival after liver transplantation for hepatitis C. *Transplantation* **77**, 84–92 (2004).
29. Arenas, J. I., Vargas, H. E. & Rakela, J. The use of hepatitis C-infected grafts in liver transplantation. *Liver Transplant.* **9**, S48–S51 (2003).
30. Saab, S. et al. Outcomes of hepatitis C and hepatitis B core antibody-positive grafts in orthotopic liver transplantation. *Liver Transplant.* **9**, 1053–1061 (2003).
31. Saab, S. et al. Hepatitis C positive grafts may be used in orthotopic liver transplantation: a matched analysis. *Am. J. Transplant.* **3**, 1167–1172 (2003).
32. Charlton, M. et al. Predictors of patient and graft survival following liver transplantation for hepatitis C. *Hepatology* **28**, 823–830 (1998).
33. Charlton, M. et al. Long-term results and modeling to predict outcomes in recipients with HCV infection: results of the NIDDK liver transplantation database. *Liver Transplant.* **10**, 1120–1130 (2004).
34. Ghobrial, R. M. et al. A 10-year experience of liver transplantation for hepatitis C: analysis of factors determining outcome in over 500 patients. *Ann. Surg.* **234**, 384–393; discussion 393–394 (2001).
35. Terrault, N. A. Hepatitis C virus and liver transplantation. *Semin. Gastrointest. Dis.* **11**, 96–114 (2000).
36. Zervos, X. A. et al. Comparison of tacrolimus with microemulsion cyclosporine as primary immunosuppression in hepatitis C patients after liver transplantation. *Transplantation* **65**, 1044–1046 (1998).
37. Wiesner, R. et al. A randomized double-blind comparative study of mycophenolate mofetil and azathioprine in combination with cyclosporine and corticosteroids in primary liver transplant recipients. *Liver Transplant.* **7**, 442–450 (2001).
38. Jain, A. et al. A prospective randomized trial of mycophenolate mofetil in liver transplant recipients with hepatitis C. *Liver Transplant.* **8**, 40–46 (2002).
39. Nelson, D. R. et al. Anti-interleukin-2 receptor therapy in combination with mycophenolate mofetil is associated with more severe hepatitis C recurrence after liver transplantation. *Liver Transplant.* **7**, 1064–1070 (2001).
40. Charlton, M. & Seaberg, E. Impact of immunosuppression and acute rejection on recurrence of hepatitis C: results of the National Institute of Diabetes and Digestive and Kidney Diseases Liver Transplantation Database. *Liver Transplant. Surg.* **5**, S107–S114 (1999).
41. Eason, J. D., Nair, S., Cohen, A. J., Blazek, J. L. & Loss, G. E. Jr. Steroid-free liver transplantation using rabbit antithymocyte globulin and early tacrolimus therapy. *Transplantation* **75**, 1396–1369 (2003).
42. Shiffman, M. L. et al. Peginterferon $\alpha 2a$ and ribavirin in patients with chronic hepatitis C who have failed prior treatment. *Gastroenterology* **126**, 1015–1023 (2004).
43. Crippin, J. S., McCashland, T., Terrault, N., Sheiner, P. & Charlton, M. R. A pilot study of the tolerability and efficacy of antiviral therapy in hepatitis C virus-infected patients awaiting liver transplantation. *Liver Transplant.* **8**, 350–355 (2002).
44. Everson, G. T. Should we treat patients with chronic hepatitis C on the waiting list? *J. Hepatol.* **42**, 456–462 (2005).
45. Everson, G. T. Treatment of chronic hepatitis C in patients with decompensated cirrhosis. *Rev. Gastroenterol. Disord.* **4** (suppl. 1), S31–S38 (2004).
46. Gane, E. Treatment of recurrent hepatitis C. *Liver Transplant.* **8**, S28–S37 (2002).
47. Ahmad, J., Dodson, S. F., Demetris, A. J., Fung, J. J. & Shakil, A. O. Recurrent hepatitis C after liver transplantation: a nonrandomized trial of interferon α alone versus interferon α and ribavirin. *Liver Transplant.* **7**, 863–869 (2001).

48. Lavezzo, B. *et al.* Treatment of recurrent hepatitis C in liver transplants: efficacy of a six versus a twelve month course of interferon α 2b with ribavirin. *J. Hepatol.* **37**, 247–252 (2002).
49. Samuel, D. *et al.* Interferon- α 2b plus ribavirin in patients with chronic hepatitis C after liver transplantation: a randomized study. *Gastroenterology* **124**, 642–650 (2003).
50. Chalasani, N. *et al.* Peginterferon α 2a for hepatitis C after liver transplantation: two randomized, controlled trials. *Hepatology* **41**, 289–298 (2005).
51. Rodriguez-Luna, H. *et al.* Treatment of recurrent hepatitis C infection after liver transplantation with combination of pegylated interferon α 2b and ribavirin: an open-label series. *Transplantation* **77**, 190–194 (2004).
52. Ross, A. S. *et al.* Pegylated interferon α 2b plus ribavirin in the treatment of post-liver transplant recurrent hepatitis C. *Clin. Transplant.* **18**, 166–173 (2004).
53. Saab, S. *et al.* Outcomes of acute rejection after interferon therapy in liver transplant recipients. *Liver Transplant.* **10**, 859–867 (2004).
54. Toniutto, P. *et al.* Pegylated versus standard interferon- α in antiviral regimens for post-transplant recurrent hepatitis C: comparison of tolerability and efficacy. *J. Gastroenterol. Hepatol.* **20**, 577–582 (2005).
55. Stravitz, R. T. *et al.* Effects of interferon treatment on liver histology and allograft rejection in patients with recurrent hepatitis C following liver transplantation. *Liver Transplant.* **10**, 850–858 (2004).
56. Wiesner, R. H., Sorrell, M. & Villamil, F. Report of the first International Liver Transplantation Society expert panel consensus conference on liver transplantation and hepatitis C. *Liver Transplant.* **9**, S1–S9 (2003).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.S.B. (rb464@columbia.edu).

Exotoxin A-eEF2 complex structure indicates ADP ribosylation by ribosome mimicry

René Jørgensen^{1*}, A. Rod Merrill^{2*}, Susan P. Yates², Victor E. Marquez⁴, Adrian L. Schwan³, Thomas Boesen¹ & Gregers R. Andersen¹

The bacteria causing diphtheria, whooping cough, cholera and other diseases secrete mono-ADP-ribosylating toxins that modify intracellular proteins. Here, we describe four structures of a catalytically active complex between a fragment of *Pseudomonas aeruginosa* exotoxin A (ETA) and its protein substrate, translation elongation factor 2 (eEF2). The target residue in eEF2, diphthamide (a modified histidine), spans across a cleft and faces the two phosphates and a ribose of the non-hydrolysable NAD⁺ analogue, β TAD. This suggests that the diphthamide is involved in triggering NAD⁺ cleavage and interacting with the proposed oxacarbenium intermediate during the nucleophilic substitution reaction, explaining the requirement of diphthamide for ADP ribosylation. Diphtheria toxin may recognize eEF2 in a manner similar to ETA. Notably, the toxin-bound β TAD phosphates mimic the phosphate backbone of two nucleotides in a conformational switch of 18S rRNA, thereby achieving universal recognition of eEF2 by ETA.

Exotoxin A belongs to a family of toxins including diphtheria toxin, pertussis toxin, cholera toxin, C3 exoenzyme and others. These toxins have four different major targets (α -subunit of heterotrimeric G proteins, actin, Rho/Rac and eEF2) in the eukaryotic cell but any of the toxins modify only a single protein or very closely related proteins. The ADP-ribosylating toxins share a common structural core that contains the NAD⁺ binding site^{1–9}. *P. aeruginosa* infections are common among immuno-compromised patients, and pronounced multi-drug resistance makes antibiotic therapy difficult¹⁰. Diphtheria was a serious childhood disease before widespread immunization, but developing countries still experience frequent outbreaks¹¹. The translation factor eEF2 catalyses the translocation of transfer RNA and messenger RNA on the ribosome. All eukaryotic and archaeal EF2 contain a single diphthamide residue, a post-translationally modified histidine, located at residue 699 in yeast eEF2 (ref. 12). Diphthamide is only found in this protein, and is the specific molecular target for ETA and diphtheria toxin, which inactivate eEF2 by transferring the ADP-ribosyl moiety of NAD⁺ onto the diphthamide imidazole (Fig. 1a; reviewed in ref. 13). The ADP ribosylation of eEF2 by ETA follows a random third-order S_N1 mechanism^{14,15}. An oxacarbenium ion intermediate with a positively charged N ribose (Fig. 1a) results from NAD⁺ cleavage by diphtheria toxin, pertussis toxin and cholera toxin¹⁶, and probably also from ETA-catalysed cleavage¹⁵.

Structure of the eEF2-ETA_c complex

We have determined four crystal structures of complexes between yeast eEF2 and a catalytically active fragment of ETA (ETA_c): (1) the *apo* complex eEF2-ETA_c; (2) the putative Michaelis complex eEF2-ETA_c- β TAD (β -methylene-thiazole-4-carboxamide adenine dinucleotide¹⁷); (3) the post-reaction complex between ADP-

ribosylated eEF2 (ADPR-eEF2) and ETA_c; and (4) the eEF2-ETA_c complex with a water-soluble ETA inhibitor, PJ34, bound within the active site¹⁸. Diffraction data extend to a maximum resolution of 2.8–3.1 Å for the four complexes (Supplementary Table 1). All structures were determined by molecular replacement with the structures of eEF2 and ETA_c. They are highly isomorphous and have three complexes in the asymmetric units (complex 1, chains A and B; complex 2, chains C and D; complex 3, chains E and F). In general, the molecules are well ordered, especially domain IV of eEF2 and ETA_c, which form most of the intermolecular contacts. Only a single β -hairpin in eEF2 domain III contributes to the interaction (Fig. 1b, c). In complex 3 eEF2 domains I, G', III and V are mobile. The conformation of eEF2 in the toxin complex is similar to that of *apo*-eEF2 (ref. 19), but owing to its interaction with ETA_c (Fig. 1b, c), eEF2 domain III has rotated 13°–18° towards domain IV.

There are no major conformational changes such as domain reorientations or large shifts in loops in either eEF2 or ETA_c when comparing the four different structures. In contrast, there are roughly two types of eEF2-ETA_c complexes in all four structures. The NAD⁺ binding sites of complexes 2 and 3 are rotated 5°–9° degrees towards the diphthamide located in domain IV of eEF2 (Fig. 1c) compared to complex 1. The axis of rotation passes through ETA residues 493–494 and 578–579 within the eEF2-toxin interface. This causes the distance between the NC1 of the N ribose and the nucleophilic NE2 from the diphthamide residue to be 11.0 Å in β TAD for complexes 2 and 3, but 12.1 Å in β TAD for complex 1. In eEF2-ETA_c- β TAD complexes 2 and 3, the diphthamide substituent reaches towards the β TAD N phosphate (Supplementary Fig. 1). The bulky and positively charged quaternary ammonium substituent on the diphthamide is oriented towards the β TAD N phosphate whereas

¹Centre for Structural Biology, Department of Molecular Biology, University of Aarhus, Gustav Wieds Vej 10C, DK-8000, Denmark. ²Department of Molecular and Cellular Biology, and ³Department of Chemistry, University of Guelph, Guelph, Ontario N1G 2W1, Canada. ⁴Laboratory of Medicinal Chemistry, Center for Cancer Research, National Cancer Institute at Frederick, NIH, Frederick, Maryland 21702, USA.

*These authors contributed equally to this work.

the amide group faces towards the N ribose. In complex 1 the diphthamide substituent is disordered.

The interfaces between eEF2 and ETA_c are slightly variable due to the different rotations of the toxin relative to eEF2 domain IV and small differences in the orientations of eEF2 domain III (Fig. 1c). Several hydrogen bonds and salt bridges are present at the intermolecular interface, but none of these is strictly conserved in all our structures. However, some key van der Waals interactions are conserved (Fig. 2b). Five residues in eEF2 domain IV consistently interact with ETA_c . Residues Trp 669 and Gly 703 interact with the ETA_c L4 region (see below), and eEF2 Arg 711, Met 828 and Tyr 838 interact with residues Arg 576 to Gly 580 in the toxin (Fig. 2b). The

three residues Trp 669, Gly 703 and Arg 711 are strictly conserved in eukaryotic eEF2, whereas Met 828 and Tyr 838 are conserved as M/L/I and F/Y¹⁹, respectively. Besides the five eEF2 residues always in contact with ETA_c , the domain III residues Met 522 to Glu 527 (where Glu 524 and Glu 527 are strictly conserved in eEF2) in the hairpin loop often interact with Arg 412, Arg 490 and Arg 492 in ETA_c .

The toxin structure and the NAD^+ pocket

The binding of yeast eEF2 to ETA_c accelerates the rate of glycosidic NC1–NN1 bond cleavage within NAD^+ by 25,000-fold²⁰, so rearrangement of ETA_c residues around the nicotinamide and N

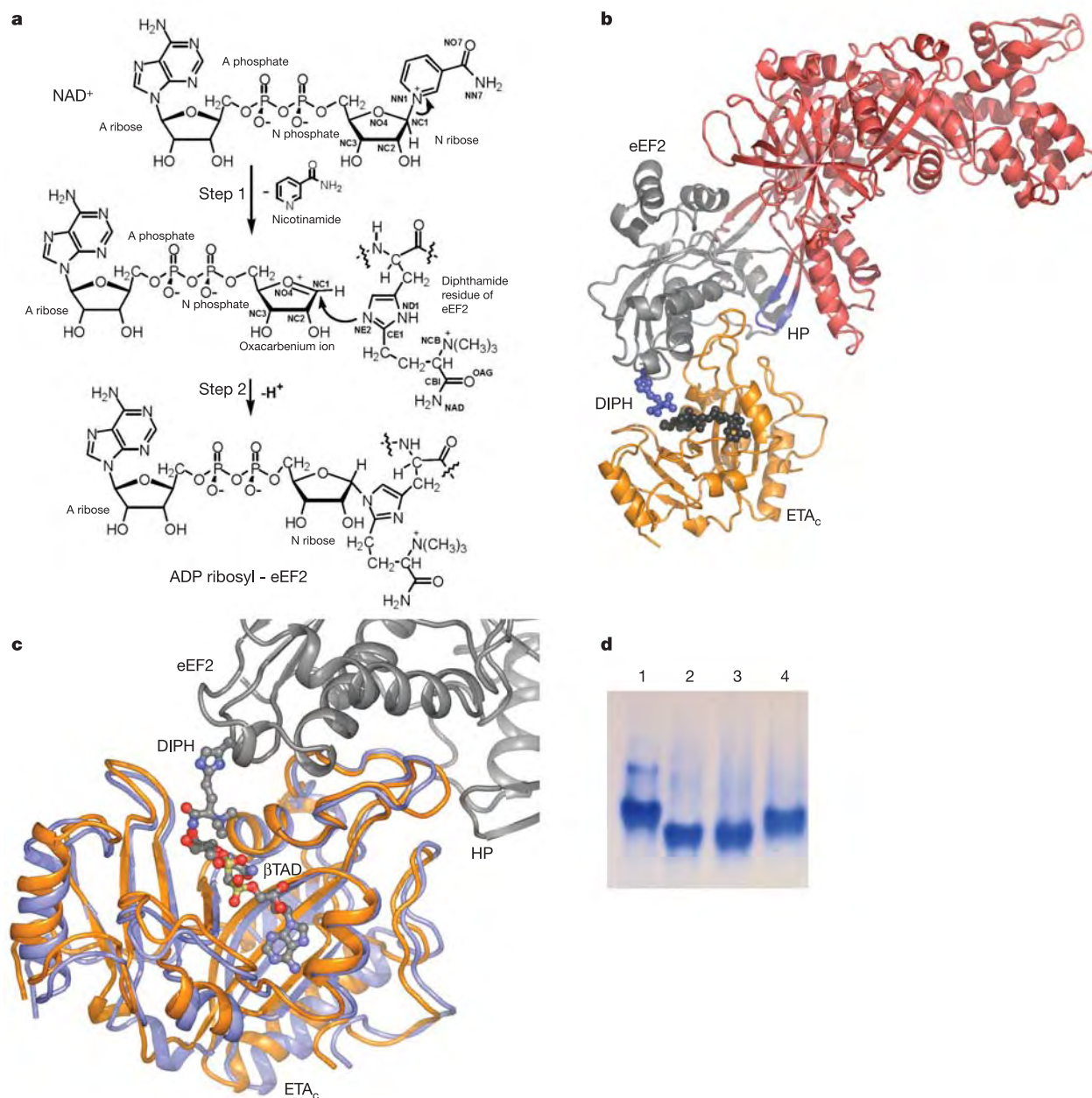


Figure 1 | The enzymatic reaction and the eEF2- ETA_c structure.

a, Schematic representation of the ADP-ribosylation reaction with important atoms labelled. The reaction features an $\text{S}_{\text{N}}1$ mechanism with scission of the glycosidic bond (NC1–NN1) within the NAD^+ substrate **b**, The toxin (orange) binds to eEF2 domain IV (grey) and a single hairpin in domain III (blue, marked HP). The diphthamide (blue, marked DIPH) and the βTAD (black) are shown in ball and stick representation. **c**, In the

eEF2- ETA_c - βTAD structure, the toxin in complex 1 (blue) is rotated 6° compared to complex 3 (orange). **d**, ADP ribosylation in the crystalline state analysed by native PAGE. Lane 1, eEF2; lane 2, ADPR-eEF2; lane 3, washed eEF2- ETA_c crystals reacted during NAD^+ soak; lane 4, washed eEF2- ETA_c crystals soaked without NAD^+ . Buffer samples used for washing the crystals either before or after the NAD^+ reaction did not contain eEF2, as analysed by PAGE (not shown).

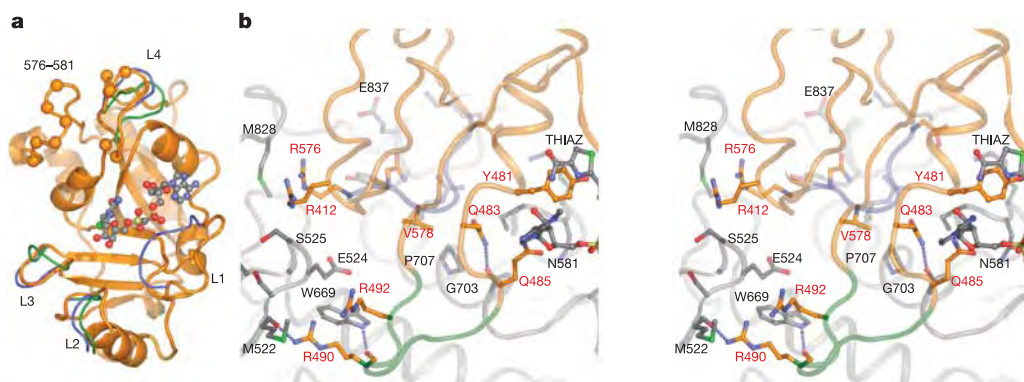


Figure 2 | The toxin loops and interface region. **a**, Structure of ETA_c (orange) in its complex with eEF2 and βTAD with the four variable loops L1–L4. Residues in the interface to eEF2 are shown with their C_α atom as a sphere. The conformation of the L1–L4 loops in the structures of the ETA mutant W281A^1 and the free ETA_c – βTAD complex² are shown in blue and green, respectively. **b**, Stereo representation of the interface found in the

eEF2– ETA_c – βTAD complex 3 (where black labels represent eEF2 and red labels represent ETA_c). The ETA_c L4 loop and residues 576–581 are shown with green and blue backbone, respectively. Three intermolecular hydrogen bonds are shown with blue dots. The thiazole ring of βTAD is marked THIAZ.

ribose might be expected after eEF2 binding, which could optimize the geometry of the toxin residues Glu 553 (ref. 21), Tyr 470 and Tyr 481 (ref. 22) around the scissile bond for bond cleavage. In previous ETA structures^{1,2,18}, four loop regions surrounding the NAD^+ pocket show structural variations: L1, 457–464; L2, 517–522; L3, 546–551 and L4, 486–493. Compared to the structure of free ETA_c , eEF2 binding induces considerable conformational changes in these loops (Fig. 2a). The L1 loop becomes fixed through a hydrogen bond between Gln 460 and Trp 558 of ETA_c but is not in direct contact with eEF2. The L2 and L3 loops are located next to each other at the edge of the NAD^+ binding pocket and are not in contact with eEF2 in the complexes (Fig. 2a). The fourth variable loop, L4, makes intimate contacts with eEF2 (Fig. 2a, b). In the interaction between eEF2 and ETA_c , L4 is the most direct linkage between eEF2 and the residues around the scissile bond in NAD^+ as it is adjacent to Tyr 481, which stacks with the NAD^+ nicotinamide moiety. The L4 loop and three adjacent residues are in contact with the hairpin loop Met 522 to Glu 527 in domain III of eEF2, residues Gly 702, Gly 703 and Ile 706 located right after the diphthamide in domain IV of eEF2, and finally Trp 669 of eEF2 (Fig. 2b).

The βTAD binding sites in our eEF2 complexes are quite similar to that of the free ETA_c – βTAD complex with only minor changes in surrounding residues. Furthermore, there are no obvious differences in the βTAD conformations. In the eEF2– ETA_c –PJ34 complex, the inhibitor is placed between the parallel walls of the nicotinamide-binding pocket as previously observed¹⁸. In contrast to the strong electron density for βTAD and PJ34 in all three complexes of the asymmetric unit, the ADP-ribose moiety can only be reliably modelled in complex 3 of the ADPR–eEF2– ETA_c complex (Fig. 3c; see also Supplementary Fig. 1). The N ribose forms hydrogen bonds with eEF2 Glu 696 and His 694, and the N phosphate interacts electrostatically with the ammonium group of the diphthamide moiety. These interactions are also observed in the structure of free ADPR–eEF2 (ref. 23), but the A ribose and the adenine base have reoriented markedly. The A ribose is close to Trp 466 and Arg 467 located between the L1 loop and Tyr 470 in ETA , and this tyrosine also stacks with the adenine base of the ADP ribose group.

The reaction mechanism

As suggested for Glu 148 in diphtheria toxin, Glu 553 of ETA is likely to stabilize the oxacarbenium intermediate after dissociation of nicotinamide by forming a hydrogen bond with the 2' OH of the N ribose², possibly aided by the phenol group of Tyr 481 (Fig. 3b). As first described for diphtheria toxin, NAD^+ bound to mono-ADP-ribosylating toxins is in a strained conformation. In particular, the

torsion angle χ_N about the scissile bond is unfavourable⁵. This strain may also explain why significant changes are not observed directly around the scissile bond upon eEF2 binding. The interaction between the diphthamide and the NAD^+ N phosphate (Fig. 3b) may help to trigger the transferase reaction, as the electrostatic interaction between the diphthamide and the N phosphate is likely to increase the strain in NAD^+ further by slightly pulling the phosphate and the ribose away from the nicotinamide. The cleavage of the NC1 – NN1 bond of NAD^+ will release the strain and generate the proposed oxacarbenium ion that will react with the NE2 of the diphthamide.

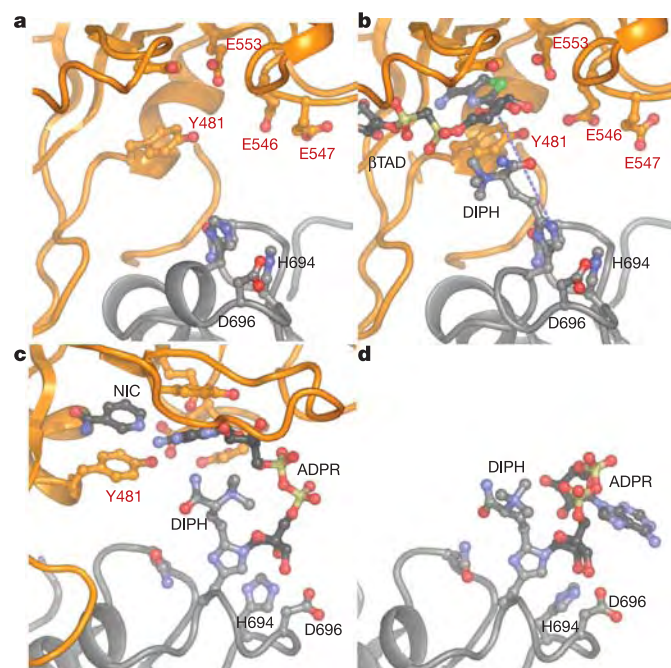


Figure 3 | Expanded view of the NAD^+ binding pocket during the reaction. **a**, The structure of the eEF2– ETA_c complex before NAD^+ binding. Colours follow the scheme of previous figures. **b**, Structure of the eEF2– ETA_c – βTAD Michaelis complex with diphthamide (DIPH) interacting with the N phosphate. The blue dotted line connects the two atoms forming the covalent bond at the end of the reaction. **c**, Structure of the post-reaction ADPR–eEF2– ETA_c complex with a docked nicotinamide. **d**, The structure of ADPR–eEF2 (ref. 23) after dissociation of ETA_c .

The large distance between the N ribose NC1 and the nucleophilic diphthamide NE2 atom observed in the Michaelis complex is surprising. To test rigorously the relevance of the structures we demonstrated that the ADP-ribosyltransferase (ADPRT) reaction readily occurs in crystals of eEF2-ETA_c soaked with NAD⁺ (Fig. 1d). These crystals are physically intact after reaction, and still diffract to at least 3.5 Å. The eEF2-ETA_c-βTAD complex is a close analogue to the enzyme-substrate (E-S) complex (Michaelis complex) for the ADPRT reaction, but it clearly does not represent the transition state complex, nor does it indicate what structural changes are required for that complex to form. There must be at least one intermediate complex in the reaction pathway, which then will lead to the formation of the transition state species for the reaction. Although we do not have any structural data on the nature of these additional complexes involved in the reaction mechanism, a governing feature of these complexes is that the distance between the ribose NC1 and the nucleophilic NE2 atom must be considerably smaller than what is observed in our Michaelis complex structure. Kinetic isotope effect studies suggest a separation between the NAD⁺ NC1 and diphthamide NE2 atoms of 2.6 Å at the transition state of the eEF2-DTA (catalytic domain of the diphtheria toxin) complex²⁴.

There are potentially several ways of reducing the nucleophile-electrophile distance to a similar value in the eEF2-ETA_c complex. One option is that the oxacarbenium ion migrates from the NAD⁺ pocket in ETA towards the diphthamide NE2 atom, but that would inevitably require that the highly reactive oxacarbenium ion was protected from reacting with water or other nucleophiles during transfer across the intermolecular cleft, which probably contains one–two layers of bound water. Presently, there is no experimental data supporting such a mechanism. The distance could also be reduced if eEF2 rotated substantially relative to ETA_c, but that seems to be prevented in the enzymatically active crystalline state by the tight crystal packing around ETA_c and eEF2 domain IV. A third possibility is that the loop region containing the diphthamide may undergo a transient conformational change during reaction, thereby bringing the diphthamide NE2 atom closer to the oxacarbenium ion. Such a conformational change has not yet been observed, but the diffraction data for all known structures of eEF2 were collected at 100 K. The diphthamide loop region may be more mobile or capable of adopting an alternative conformation at room temperature, where the complex is enzymatically active in the crystal. The interaction between the diphthamide ammonium group and the NAD⁺ phosphates may promote a conformational change in the loop. Furthermore, the relaxation process associated with cleavage of the

NC1-NN1 bond of NAD⁺ and alleviation of the strained conformation may also contribute to partial reduction of the distance between the oxacarbenium ion and the diphthamide without exposing the positive charge. However, no matter how the nucleophile-electrophile distance is reduced it seems likely that the quaternary ammonium of the diphthamide modification remains directed towards the phosphates of NAD⁺, thus anchoring the oxacarbenium ion to the diphthamide (Fig. 3).

Our structures and a reaction mechanism that features a relatively long distance between the initial positions of NAD⁺ NC1 and diphthamide NE2—which requires additional steps for completion—is consistent with important experimental observations. First, the reaction mechanism is random third order¹⁴, which can be explained by the large cleft between eEF2 and ETA_c at the NAD⁺ binding site, allowing the nucleotide to diffuse in after complex formation between eEF2 and ETA_c—a property nicely illustrated by our ability to diffuse NAD⁺ into crystals of eEF2-ETA_c (Fig. 3a) and observe transfer of the ADP-ribosyl moiety (Fig. 1d). Second, the interaction of the diphthamide modification with the βTAD phosphates explains why the diphthamide is required for ADP ribosylation²⁵. Third, the gap between the electrophile and the nucleophile supports the proposed S_N1 mechanism for ADP ribosylation by ETA and implies a multi-step process¹⁵. Fourth, except for Asn 581 of eEF2, which is tightly engaged in a hydrogen bond with eEF2 Gln 704 and the amide group of the diphthamide moiety, the two nitrogen atoms of the diphthamide imidazole ring and one nitrogen in eEF2 His583 are the closest prospective nucleophilic atoms in eEF2, with approximately the same distance to the βTAD NC1 atom. The suggested tight association of the oxacarbenium ion with the diphthamide moiety is likely to explain the preference for the diphthamide NE2 atom. Fifth, the structures confirm previous mutational studies identifying L4 of ETA as being much more important for ADP ribosylation than L1 (ref. 26). Finally, the eEF2-ETA cleft appears to mimic a cleft of similar dimensions between eEF2 and the 80S ribosome (see below).

Implications for ETA_f and other toxins

ETA enters the cell by receptor-mediated endocytosis²⁷, and the catalytic carboxy-terminal fragment (ETA_f) is generated by furin cleavage at Gly 280 and reduction of the Cys 265–Cys 287 disulphide bridge (ref. 28). We superimposed a model of ETA_f derived directly from the structure of full-length ETA¹ onto the ETA_c from our eEF2 complexes (Fig. 4a). This demonstrates that binding of ETA_f to eEF2 does not require substantial conformational changes in ETA_f compared to the crystal structure of full-length ETA. However, the well-defined L1 conformation observed in our structures of eEF2-ETA_c

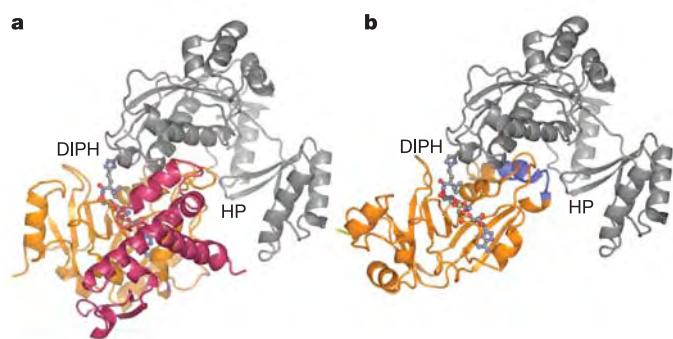


Figure 4 | Models of eEF2-toxin complexes based on the eEF2-ETA_c-βTAD complex. **a**, A model of ETA_f was derived from the structure of ETA¹. Residues 281–398 of ETA are coloured dark red. **b**, Model of DTA³ in complex with eEF2. Residues 170–180 in diphtheria toxin are coloured blue. The eEF2 diphthamide and βTAD are shown in ball and stick representation. The interacting β-hairpin from eEF2 domain III and the diphthamide are labelled HP and DIPH, respectively.

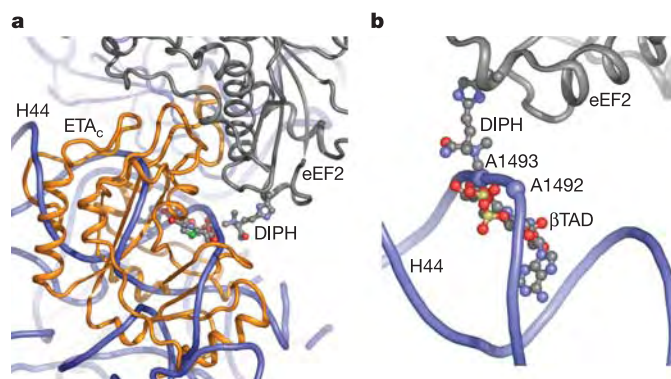


Figure 5 | Ribosome mimicry by ETA. **a**, The toxin (orange) and the large helix 44 (H44) in the 40S ribosomal subunit (blue) share overall localization with respect to domain IV of eEF2 (grey). **b**, The phosphates of βTAD (NAD⁺) are presented to the eEF2 diphthamide in an orientation markedly similar to the sugar-phosphate backbone of bases A1492 and A1493.

(Fig. 2a) would cause a clash with residues 341–347 in ETA_c . This suggests that either the L1 loop is shifted towards eEF2 in the physiological complex, or residues 280–398 in ETA_c rearrange slightly to accommodate the L1 conformation observed in our eEF2– ETA_c complexes.

A superposition of the catalytic domain of diphtheria toxin (DTA)³ onto ETA_c shows that DTA may interact with eEF2 in the same manner as ETA_c (Fig. 4b). Only modest rearrangement of the flexible L4 loop is required. The model predicts DTA to interact with accessible residues within regions 701–710 and 665–673 of eEF2 domain IV. The Arg 576 to Asp 581 segment in ETA_c is also part of the eEF2–toxin interface (Fig. 2). Instead of a loop in ETA_c , DTA has an α -helix at this location (Fig. 4b). However, contacts to eEF2 still appear quite feasible between solvent-exposed residues within DTA Arg 170 to Glu 180 and eEF2 residues 833–838 in domain IV or 523–526 in domain III.

From our structures of the eEF2– ETA_c – βTAD and ADPR–eEF2– ETA_c complexes, it is clear that residues of eEF2 are important for the reaction mechanism; hence the reaction is substrate-assisted. The diphthamide ammonium group is located close to the N phosphate both before and after the ADPRT reaction, and Asp 696 forms a hydrogen bond with the N ribose 2' OH in ADPR–eEF2. As in the eEF2– ETA_c system, ADP ribosylation by other toxins might also be substrate-assisted. Diphthamide is unique to eEF2, whereas the acceptor residue is different for other toxin targets: for example, an arginine in $G_{\alpha s}$ and actin, cysteine in $G_{\alpha i}$, (where s and i indicate stimulatory and inhibitory α -subunits of heterotrimeric G proteins, respectively) and asparagine in Rho²⁹. The diphthamide modification may be functionally homologous to arginine or lysine in other toxin targets, whereas an aspartate or a glutamate could be equivalent to eEF2 Asp 696. Analysis of the region neighbouring the acceptor residues of four different ADP-ribosylation targets showed that a highly conserved, surface-exposed arginine and aspartate/glutamate could be identified that might be analogous to the eEF2 diphthamide and Asp 696 (Supplementary Fig. 2). We suggest that the putative Asp 696 analogues possibly interact with the N ribose hydroxyl groups during or after the ADPRT reaction. The conserved, exposed arginines may interact with the phosphates in NAD^+ during catalysis or in the ADP-ribosylated target residue after catalysis.

Ribosome mimicry ensures recognition

In the cryo-electron microscopy reconstruction of eEF2 in complex with the 80S ribosome³⁰, the area around the diphthamide appears to interact with the decoding site in the 40S ribosomal subunit. To investigate whether ETA_c takes advantage of this fundamental property of eEF2 to obtain universal recognition of eukaryotic and archaeal EF2, we superimposed eEF2 domain IV from the βTAD –toxin complex onto domain IV of eEF2 from the eEF2–80S complex (Fig. 5a). This reveals a striking similarity between the overall orientation of the toxin and helix 44 from 18S rRNA with respect to eEF2. In particular, the open cleft between eEF2 and the toxin NAD^+ binding site mirrors the distance between helix 44 and eEF2. Even more spectacular is the mimicry made by the phosphates of βTAD (Fig. 5b), which coincide almost perfectly with the backbone phosphates of A1492 and A1493 (*Escherichia coli* numbering) in helix 44. These two adenines are universally conserved in 16S/18S rRNA, and are essential for tRNA recognition at the A site, as they monitor a cognate codon–anticodon helix³¹, thereby discriminating effectively against binding of near-cognate tRNA.

By optimizing the mimicry of this essential piece of rRNA during evolution, *Pseudomonas aeruginosa* exotoxin A minimizes the probability that the target organism could evolve resistance towards the invading toxin, as this would require coordinated mutations in regions of eEF2 and the ribosome that are crucial for function. One further implication may be that part of the normal function of diphthamide on the ribosome is to contact the backbone around A1493 (Fig. 5b) at the end of translocation. The two bases, A1492 and

A1493, alternate between being stacked in an internal loop within helix 44 in the absence of tRNA in the A site, and flipped out in the presence of cognate tRNA (Supplementary Fig. 3), so this conformational switch must be reset after each elongation cycle. eEF2 and especially the diphthamide might stabilize the stacked conformation of the two bases. Finally, the mimicry of these two rRNA nucleotides by the toxin– NAD^+ complex is also an overwhelming piece of evidence for the *in vivo* relevance of our structures of the eEF2– ETA_c complex.

METHODS

Purification of yeast eEF2, ADPR–eEF2 and *E. coli* expressed ETA_c was done as described^{23,32,33}. Purified ETA_c protein in 20 mM Tris–HCl, pH 7.6, 100 mM NaCl was concentrated to 10 mg ml^{−1}, and eEF2 or ADPR–eEF2 in 75 mM KCl, 5 mM MgCl₂, 0.5 mM EDTA, 0.5 mM dithiothreitol (DTT) and 20 mM HEPES, pH 7.2 was concentrated to 7–9 mg ml^{−1}. ETA_c was mixed with eEF2 or ADPR–eEF2 in a 7:1 molar ratio in buffer A (20 mM HEPES, 75 mM KCl, 5 mM MgCl₂, 0.5 mM EDTA, 0.5 mM DTT). For the ligand complexes either 5 mM βTAD or 0.5 mM PJ34 (Sigma–Aldrich) was added in all subsequent steps. The mixture was then precipitated in 15% PEG 8K. The resulting pellet was washed in buffer A containing 15% PEG 8K and then gently re-suspended in buffer A. Crystallization was done by vapour diffusion against reservoirs containing 10% PEG 6K, 5% 2-methyl-2,4-pentanediol (MPD), 3 mM DTT, 0.9 mM Na₃N and 100 mM HEPES, pH 7.2 at 19°C. Crystals were transferred to cryo-protection buffer (10% PEG 6K, 5% MPD, 3 mM DTT, 0.9 mM Na₃N and 100 mM HEPES, pH 7.2 in 8.5% glycerol) and flash frozen in liquid N₂. For the eEF2– ETA_c –PJ34 complex an initial 3.2 Å data set was collected at beamline BL14.1. This structure was determined by molecular replacement with MOL-REP³⁴ using eEF2 domains I–II (RCSB protein data bank entry 1N0V), eEF2 domains III–V, and the structure of ETA_c (RCSB entry 1AER) as search models. The model was rebuilt with program O³⁵ and refined with CNS³⁶ to an R_{free} of 29% under tight NCS restraints applied domain wise. This model was then used for rigid body refinement against the high-resolution PJ34 complex data (Supplementary Table 1) after extension and transfer of the test set. The model was then iteratively rebuilt and refined at 2.8 Å resolution. The final model of the PJ34 complex was then used for molecular replacement with the three other structures. The rotation of ETA_c relative to eEF2 and conformational changes in eEF2 were analysed with DYNDOM³⁷. Figures were prepared with PYMOL³⁸. For assaying enzymatic activity in the crystalline state, ETA_c –eEF2 crystals were washed with 20 μ l of crystallization reservoir solution three times, followed by re-suspension in 10 μ l of reservoir solution. One microlitre of 5 mM NAD^+ dissolved in reservoir solution was added to the drop containing the crystals and the ADPRT reaction was allowed to proceed for 30 min at 20°C. The reaction was stopped by washing the crystals again (three times) with 20 μ l of reservoir solution followed by dissolution of the crystals in Laemmli sample buffer without SDS. The samples were run on a 6% native acrylamide gel (Tris–glycine, pH 8.1, no SDS) containing 0.1% Triton-X100 where the eEF2– ETA_c complex dissolves completely and ADPR–eEF2 migrates faster than eEF2.

Received 9 March; accepted 3 June 2005.

1. Wedekind, J. E. *et al.* Refined crystallographic structure of *Pseudomonas aeruginosa* exotoxin A and its implications for the molecular mechanism of toxicity. *J. Mol. Biol.* **314**, 823–837 (2001).
2. Li, M., Dyda, F., Benhar, I., Pastan, I. & Davies, D. R. Crystal structure of the catalytic domain of *Pseudomonas* exotoxin A complexed with a nicotinamide adenine dinucleotide analog: implications for the activation process and for ADP ribosylation. *Proc. Natl Acad. Sci. USA* **93**, 6902–6906 (1996).
3. Weiss, M. S., Blanke, S. R., Collier, R. J. & Eisenberg, D. Structure of the isolated catalytic domain of diphtheria toxin. *Biochemistry* **34**, 773–781 (1995).
4. Bell, C. E. & Eisenberg, D. Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Biochemistry* **35**, 1137–1149 (1996).
5. Bell, C. E., Yeates, T. O. & Eisenberg, D. Unusual conformation of nicotinamide adenine dinucleotide (NAD) bound to diphtheria toxin: a comparison with NAD bound to the oxidoreductase enzymes. *Protein Sci.* **6**, 2084–2096 (1997).
6. Han, S., Arvai, A. S., Clancy, S. B. & Tainer, J. A. Crystal structure and novel recognition motif of rho ADP-ribosylating C3 exoenzyme from *Clostridium botulinum*: structural insights for recognition specificity and catalysis. *J. Mol. Biol.* **305**, 95–107 (2001).
7. Menetrey, J. *et al.* NAD binding induces conformational changes in Rho ADP-ribosylating *Clostridium botulinum* C3 exoenzyme. *J. Biol. Chem.* **277**, 30950–30957 (2002).
8. Evans, H. R. *et al.* The crystal structure of C3stau2 from *Staphylococcus aureus* and its complex with NAD. *J. Biol. Chem.* **278**, 45924–45930 (2003).

9. Han, S., Craig, J. A., Putnam, C. D., Carozzi, N. B. & Tainer, J. A. Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex. *Nature Struct. Biol.* **6**, 932–936 (1999).
10. Lyczak, J. B., Cannon, C. L. & Pier, G. B. Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect.* **2**, 1051–1060 (2000).
11. Mattos-Guaraldi, A. L., Moreira, L. O., Damasco, P. V. & Hirata Junior, R. Diphtheria remains a threat to health in the developing world—an overview. *Mem. Inst. Oswaldo Cruz* **98**, 987–993 (2003).
12. Van Ness, B. G., Howard, J. B. & Bodley, J. W. ADP-ribosylation of elongation factor 2 by diphtheria toxin. Isolation and properties of the novel ribosyl-amino acid and its hydrolysis products. *J. Biol. Chem.* **255**, 10717–10720 (1980).
13. Collier, R. J. Understanding the mode of action of diphtheria toxin: a perspective on progress during the 20th century. *Toxicon* **39**, 1793–1803 (2001).
14. Armstrong, S. & Merrill, A. R. Toward the elucidation of the catalytic mechanism of the mono-ADP-ribosyltransferase activity of *Pseudomonas aeruginosa* exotoxin A. *Biochemistry* **43**, 183–194 (2004).
15. Beattie, B. K., Prentice, G. A. & Merrill, A. R. Investigation into the catalytic role for the tryptophan residues within domain III of *Pseudomonas aeruginosa* exotoxin A. *Biochemistry* **35**, 15134–15142 (1996).
16. Zhou, G. C. *et al.* Inhibitors of ADP-ribosylating bacterial toxins based on oxacarbenium ion character at their transition states. *J. Am. Chem. Soc.* **126**, 5690–5698 (2004).
17. Marquez, V. E. *et al.* Thiazole-4-carboxamide adenine dinucleotide (TAD). Analogues stable to phosphodiesterase hydrolysis. *J. Med. Chem.* **29**, 1726–1731 (1986).
18. Yates, S. P. *et al.* Structure-function analysis of water soluble inhibitors of the catalytic domain of exotoxin A from *Pseudomonas aeruginosa*. *Biochem. J.* **385**, 667–675 (2004).
19. Jorgensen, R. *et al.* Two crystal structures demonstrate very large conformational changes of the eukaryotic ribosomal translocase. *Nature Struct. Biol.* **10**, 379–385 (2003).
20. Yates, S. P. & Merrill, A. R. Characterization of oxidized nicotinamide adenine dinucleotide (NAD⁺) analogues using a high-pressure-liquid-chromatography-based NAD⁺-glycohydrolase assay and comparison with fluorescence-based measurements. *Anal. Biochem.* **340**, 41–51 (2005).
21. Douglas, C. M. & Collier, R. J. Exotoxin A of *Pseudomonas aeruginosa*: substitution of glutamic acid 553 with aspartic acid drastically reduces toxicity and enzymatic activity. *J. Bacteriol.* **169**, 4967–4971 (1987).
22. Lukac, M. & Collier, R. J. *Pseudomonas aeruginosa* exotoxin A: effects of mutating tyrosine-470 and tyrosine-481 to phenylalanine. *Biochemistry* **27**, 7629–7632 (1988).
23. Jorgensen, R. *et al.* Crystal structure of ADP-ribosylated ribosomal translocase from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**, 45919–45925 (2004).
24. Parikh, S. L. & Schramm, V. L. Transition state structure for ADP-ribosylation of eukaryotic elongation factor 2 catalyzed by diphtheria toxin. *Biochemistry* **43**, 1204–1212 (2004).
25. Liu, S. & Leppla, S. H. Retroviral insertional mutagenesis identifies a small protein required for synthesis of diphthamide, the target of bacterial ADP-ribosylating toxins. *Mol. Cell* **12**, 603–613 (2003).
26. Yates, S. P. & Merrill, A. R. A catalytic loop within *Pseudomonas aeruginosa* exotoxin A modulates its transferase activity. *J. Biol. Chem.* **276**, 35029–35036 (2001).
27. Kounnas, M. Z. *et al.* The alpha 2-macroglobulin receptor/low density lipoprotein receptor-related protein binds and internalizes *Pseudomonas* exotoxin A. *J. Biol. Chem.* **267**, 12420–12423 (1992).
28. Chiron, M. F., Fryling, C. M. & FitzGerald, D. J. Cleavage of pseudomonas exotoxin and diphtheria toxin by a furin-like enzyme prepared from beef liver. *J. Biol. Chem.* **269**, 18167–18176 (1994).
29. Corda, D. & Di Girolamo, M. Functional aspects of protein mono-ADP-ribosylation. *EMBO J.* **22**, 1953–1958 (2003).
30. Spahn, C. M. *et al.* Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J.* **23**, 1008–1019 (2004).
31. Ogle, J. M. *et al.* Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902 (2001).
32. Jorgensen, R., Carr-Schmid, A., Ortiz, P. A., Kinzy, T. G. & Andersen, G. R. Purification and crystallization of the yeast elongation factor eEF2. *Acta Crystallogr. D* **58**, 712–715 (2002).
33. Armstrong, S., Yates, S. P. & Merrill, A. R. Insight into the catalytic mechanism of *Pseudomonas aeruginosa* exotoxin A. Studies of toxin interaction with eukaryotic elongation factor-2. *J. Biol. Chem.* **277**, 46669–46675 (2002).
34. Vagin, A. & Teplyakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D* **56**, 1622–1624 (2000).
35. Jones, T. A., Cowan, S., Zou, J.-Y. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
36. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
37. Hayward, S. & Berendsen, H. J. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30**, 144–154 (1998).
38. DeLano, W. L. *The PyMOL User's Manual* (DeLano Scientific, San Carlos, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements G.R.A. was supported by the Benzon Foundation, HFSP, EU FP5 and SNF. A.R.M. was supported by CIHR and CCFF. We are grateful to R. Søe for purification of eEF2, and to J. Nyborg and D. Brodersen for reading of the manuscript.

Author Contributions R.J., A.R.M., V.E.M. and G.R.A. performed the experimental work, whereas S.P.Y., A.L.S. and T.B. analysed data.

Author Information Coordinates and structure factors are deposited at the RCSB data bank as entries 1ZM2, 1ZM3, 1ZM4 and 1ZM9. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.R.A. (gra@mb.au.dk) or A.R.M. (rmerrill@uoguelph.ca).

An unexpectedly rapid decline in the X-ray afterglow emission of long γ -ray bursts

G. Tagliaferri¹, M. Goad², G. Chincarini^{1,3}, A. Moretti¹, S. Campana¹, D. N. Burrows⁴, M. Perri⁵, S. D. Barthelmy⁶, N. Gehrels⁶, H. Krimm^{6,7}, T. Sakamoto^{6,8}, P. Kumar⁹, P. I. Mészáros⁴, S. Kobayashi⁴, B. Zhang¹⁰, L. Angelini^{6,11}, P. Banat¹, A. P. Beardmore², M. Capalbi⁵, S. Covino¹, G. Cusumano¹², P. Giommi⁵, O. Godet², J. E. Hill⁴, J. A. Kennea⁴, V. Mangano¹², D. C. Morris⁴, J. A. Nousek⁴, P. T. O'Brien², J. P. Osborne², C. Pagani^{1,4}, K. L. Page², P. Romano¹, L. Stella¹³ & A. Wells²

'Long' γ -ray bursts (GRBs) are commonly accepted to originate in the explosion of particularly massive stars, which give rise to highly relativistic jets. Inhomogeneities in the expanding flow result in internal shock waves that are believed to produce the γ -rays we see^{1,2}. As the jet travels further outward into the surrounding circumstellar medium, 'external' shocks create the afterglow emission seen in the X-ray, optical and radio bands^{1,2}. Here we report observations of the early phases of the X-ray emission of five GRBs. Their X-ray light curves are characterised by a surprisingly rapid fall-off for the first few hundred seconds, followed by a less rapid decline lasting several hours. This steep decline, together with detailed spectral properties of two particular bursts, shows that violent shock interactions take place in the early jet outflows.

The GRB prompt γ -ray emission usually goes through a strong spectral evolution, with the peak of the emission rising to higher energies in the early phases and then moving to lower energies³, whereas the subsequent afterglow phase has an X-ray spectrum that is well represented by a power law model with an energy index of ~ 1 (see, for example ref. 4). The transition from the prompt γ -ray to the afterglow emission is expected to occur in the first few minutes following a GRB (see ref. 5 and references therein). Multiwavelength observations of this transition and the early afterglow emission provide very important information regarding the properties and composition of the material released in these explosions, thus providing insight into the nature of the central engine⁶. Until now, however, this crucial time interval was largely unexplored (a few GRB afterglows at early times were observed, but with limited statistics^{7–10}, while for a few other bursts a very early optical emission was detected (see, for example, refs 11, 12). With the successful launch of the Swift¹³ satellite in November 2004, the situation has dramatically improved. We are now able to study this early afterglow phase starting a few tens of seconds after the burst explosion^{14,15}.

For five of the first seven GRBs promptly repositioned by Swift (to place each GRB inside the field of view of the two narrow-field instruments), the X-ray light curve, as seen by the X-Ray Telescope¹⁶ (XRT) on board Swift, faded very fast (see Fig. 1 and its legend). GRB050126 and GRB050219a are the first two bursts with an X-ray light curve well sampled by XRT, allowing us a detailed investigation of their properties. They were detected and located by the Burst Alert Telescope¹⁷ (BAT) on board Swift on 26 January and

19 February 2005, respectively^{18,19}. In both cases Swift promptly slewed to the BAT burst locations and the XRT immediately began taking data, detecting bright and rapidly fading X-ray counterparts. In Fig. 2 we plot the BAT 20–150 keV light curve of these two bursts and the very early phases of the associated X-ray sources seen by XRT in the 0.2–10 keV band. The most striking feature of these X-ray light curves is their very steep initial decline, followed by a flattening a few hundred seconds later which is well represented by a broken power law model (see Table 1). The light curve of GRB050219a is well sampled, and besides the general decay it clearly shows rapid variability on a timescale of a few tens of seconds that in any case does not affect the general trend. In the following, for simplicity, we will refer to these X-ray sources as the afterglows, though we note that the early X-ray emission may instead be associated with the prompt emission from the burst. We will return to this subject later.

We sought a possible delay of the afterglow onset by fitting the two X-ray light curves with a single power law model $\propto (t - t_0)^{-\alpha}$ (where t is time and t_0 would be the onset of the afterglow). In both cases, the decaying light curves can be fitted if the onset of the afterglow is shifted to $t_0 \approx 100$ s after the burst trigger with a power law slope of ~ 1 , as is typical of previously observed afterglows. For both GRBs (for $t_0 > 80$ s) the decay index in the first few hundred seconds would be $\alpha \leq 1.5$ and the emission is consistent with synchrotron radiation in the forward shock. In this case the spectral index β and the temporal index α ($f_\nu \propto \nu^{-\beta} t^{-\alpha}$) (where ν is the frequency and f_ν is the flux, in $\text{erg cm}^{-2} \text{s}^{-1} \text{Hz}^{-1}$) must obey the relation $\beta = p/2$ and $\alpha = (3\beta - 1)/2$ (ref. 6, assuming that the cooling frequency is below 0.2 keV), which is indeed satisfied by these bursts ($p = 2.7 \pm 0.6$ for GRB050126 and $p = 2.2 \pm 0.3$ for GRB050219a). However, whereas in the case of GRB050126 the light curve does not allow us to state clearly if a $(t - t_0)$ model is better than a broken power law model, for GRB050219a a broken power law definitely provides a better fit. Moreover, for this burst we detected the X-ray afterglow emission at least as early as 87 s from the trigger (see Fig. 2 and its legend). Thus, although the maximum of the afterglow emission could be at 105 s, the afterglow onset is clearly occurring before. Although an onset of the afterglow some time after the trigger can be expected, for GRB050219a our data are not consistent with a single power law for the early afterglow decay, whenever it starts (t_0).

¹INAF – Osservatorio Astronomico di Brera, Via Bianchi 46, I-23807 Merate, Italy. ²Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK.

³Università degli studi di Milano-Bicocca, Piazza delle Scienze 3, I-20126 Milano, Italy. ⁴Department of Astronomy & Astrophysics, Pennsylvania State University, Pennsylvania 16802, USA. ⁵ASI Science Data Center, Via Galileo Galilei, I-00044 Frascati, Italy. ⁶NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ⁷Universities Space Research Association, 10227 Wincopin Circle, Suite 212, Columbia, Maryland 21044, USA. ⁸National Research Council, 2101 Constitution Avenue, NW, Washington DC 20418, USA. ⁹Department of Astronomy, University of Texas, RLM 15.308, Austin, Texas 78712-1083, USA. ¹⁰Department of Physics, University of Nevada, Box 454002, Las Vegas, Nevada 891, USA. ¹¹Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA. ¹²INAF – Istituto di Astrofisica Spaziale e Cosmica. Via Ugo La Malfa 153, I-90146 Palermo, Italy. ¹³INAF-Osservatorio Astronomico di Roma, Via di Frascati 33, I-00040 Monteporzio, Italy.

To investigate whether the prompt and the early afterglow emissions are related, we converted both the BAT and XRT count rates to flux in the 0.2–10 keV band using the conversion factors derived from the BAT and XRT spectral analyses (see Table 2 and Fig. 3). As shown in Fig. 3, for GRB050219a the BAT and XRT light curves are discontinuous, whereas for GRB050126, although unlikely, this could still be possible. For GRB050219a, the afterglow emission peaks about 100 s after the burst trigger and this occurs after the prompt emission has already faded away. The spectral index during the burst for these GRBs is quite different from the spectrum during the early X-ray emission. However, given that burst spectra can go through strong spectral evolution, the XRT spectrum at the beginning could still be due, at least partly, to the prompt emission. Thus, we search for a possible spectral evolution of the X-ray sources detected by XRT. There is no significant evidence in either source for an X-ray spectral change across the break (with some caution for GRB050126, see Table 1). Thus, given that the BAT and XRT spectra are very different, we have a clear indication, at least for GRB050219a,

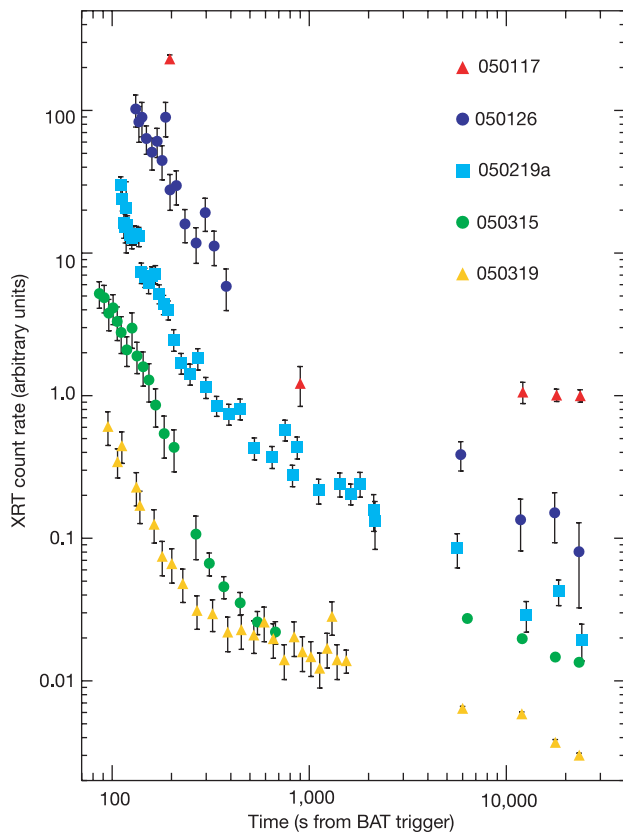


Figure 1 | The steep early X-ray light curves of five GRBs observed by XRT¹⁶. During the Swift¹³ performance verification phase, which ended on 5 April 2005, seven GRBs, discovered by the BAT¹⁷ on board Swift, were promptly repositioned by the satellite. The XRT began taking data starting up to a few tens of seconds after the burst explosion. A bright and fading X-ray counterpart was always detected. For five of them, whose light curves are shown in this figure, the fading was very fast, flattening after a few hundred seconds. For each GRB the XRT count rates are rescaled by an arbitrary constant factor for clarity, while the error bars represent the standard deviation. GRB050126 and GRB050219a are the first two bursts with X-ray light curves well sampled by XRT. The UV-Optical Telescope²⁷ (UVOT) on board Swift could not observe the field of GRB050126, owing to the proximity of the bright star Vega. Four and a half hours after the burst a new infrared source was detected in the Ks band by the Keck telescope within the XRT position error circle, with a subsequent redshift determination $z = 1.29$ of the host galaxy²⁸. For GRB050219a the UVOT did not find an optical counterpart to the X-ray source down to a limiting magnitude of $V = 20.7$ (ref. 29). No optical/near-infrared or radio counterpart to this GRB has been reported. The error bars along the y axis are 1 σ .

Table 1 | BAT and XRT best fit parameters

	GRB050126	GRB050219a
BAT time averaged spectral parameters		
T_{90} (s)	25.7	23.6
Fluence*	$(1.7 \pm 0.3) \times 10^{-6}$	$(5.2 \pm 0.4) \times 10^{-6}$
Model	Power law	Cut-off power law
β	0.34 ± 0.14	-0.75 ± 0.30
E_{peak} (keV)		90 ± 9
$\chi^2_{\text{red}}/\text{d.o.f.}$	1.25/57	0.86/56
XRT first-orbit spectral fit and parameters		
$N_{\text{H,Gal}}$ (cm^{-2})	5.3×10^{20}	8.5×10^{20}
$N_{\text{H,excess}}$ (cm^{-2})	—	$2.25 \pm 0.60 \times 10^{21}$
β	1.26 ± 0.22	1.1 ± 0.2
$\chi^2_{\text{red}}/\text{d.o.f.}$	1.06/8	1.02/53
XRT light curve fits		
α_1	$2.52^{+0.50}_{-0.22}$	3.15 ± 0.22
Break (s)	425^{+560}_{-120}	307 ± 26
α_2	$1.00^{+0.17}_{-0.26}$	0.82 ± 0.07
C-statistics/n.d.p.	26.1/20	
$\chi^2_{\text{red}}/\text{d.o.f.}$		1.41/39
t_0 (s)	105^{+9}_{-11}	105 ± 5
α	1.08 ± 0.09	0.98 ± 0.05
C-statistics/n.d.p.	31.7/20	
$\chi^2_{\text{red}}/\text{d.o.f.}$		2.6/40

T_{90} , the time interval during which 90% of the total observed counts have been detected. The start of this interval is defined by the time at which 5% of the total counts had been detected, and the end by the time at which 95% of the total counts have been detected. $N_{\text{H,Gal}}$, the hydrogen column density along the line of sight in our Galaxy. $N_{\text{H,excess}}$, the hydrogen column density in excess of the Galactic value (that is, the GRB host galaxy). d.o.f., degrees of freedom. n.d.p., number of data points. The BAT average spectrum of GRB050126 is well fitted by a simple power law ($f_E \propto E^{-\beta}$); whereas for GRB050219a, a cut-off power law ($f_E \propto E^{-\beta} e^{-E/E_0}$) is necessary. Moreover, for the latter we detect clear spectral evolution (see Table 2). For both GRBs, the XRT spectral analysis is performed on the spectra accumulated over the first orbit. We also checked for a spectral evolution across the break. In both cases, a combined fit to the two spectra accumulated before and after the break with a single absorbed power law model provides a good fit with a value for the reduced χ^2 of $\chi^2_{\text{red}} \approx 1$. If we fit the two spectra separately, leaving also the N_{H} free to vary for GRB050219a, we have an indication that the spectrum becomes harder after the break ($\alpha \approx 1.7$ for both GRBs), but the slopes are fully consistent at the 90% confidence range. Moreover, if we tied the N_{H} to be the same in the fit of the two X-ray spectra associated with GRB050219a, the two spectral indexes differ by less than 0.1. Thus, for this GRB we can definitively say that there is no spectral evolution across the break. In contrast, for GRB050126 the statistic after the break is such that we cannot speculate much about its X-ray spectrum. For this source we would be able to see a spectral change only if $\Delta\alpha \geq 0.8$. The XRT light curves shown in Fig. 1, with the onset time coinciding with the onset of the prompt, cannot be fitted by a simple power law model; an F -test shows that the chance probability for the improvement of the broken power law model is less than 10^{-4} and 10^{-8} for GRB050126 and GRB050219a, respectively. We also report the results of the best fit for a single power law model where also the onset time (t_0) is fitted. All errors quoted in the Table are for a $\Delta\chi^2 = 2.71$. For the fit of the GRB050126 light curve we used the C-statistics, because of the small number of counts in each bin. *erg cm^{-2} in the band 15–350 keV.

that the XRT source is due only to the afterglow emission. This property, together with the discontinuity in the light curve, suggests that the burst and the early afterglow emission are produced by different mechanisms. This conclusion is in agreement with the expectation that the prompt γ -ray radiation is produced in internal shocks whereas the afterglow radiation is produced in the external shock^{6,20}.

The prompt γ -ray and the early X-ray afterglow emission of GRB050219a and probably also of GRB050126 require at least two

Table 2 | GRB050219a BAT spectral fits

	Int. 1 (8 s)	Int. 2 (6 s)	Int. 3 (4 s)	Int. 4 (7 s)	Int. 4 + 5 (14 s)
β	$-2.0^{+1.0}_{-0.7}$	$-1.35^{+0.65}_{-0.55}$	$-0.70^{+0.40}_{-0.35}$	$-0.65^{+0.45}_{-0.40}$	$-0.35^{+0.50}_{-0.40}$
E_{peak}	84^{+9}_{-13}	87^{+10}_{-15}	127^{+22}_{-50}	73^{+7}_{-12}	68^{+7}_{-13}
$\chi^2_{\text{red}}/\text{d.o.f.}$	1.01/56	0.77/56	0.92/56	0.69/56	0.94/56

BAT best fit parameters for the prompt spectra of GRB050219a. For this burst we find clear evidence of spectral evolution in the prompt emission light curve, and perform the spectral analysis over five consecutive time intervals (see Fig. 2). Spectra for the first four intervals are well fitted by a cut-off power law that shows a hardening up to interval three and then a softening. The fifth interval does not have enough counts to constrain the parameters. Here we report the time duration and the spectral fit parameters for the first four intervals and for the fourth and fifth intervals added together.

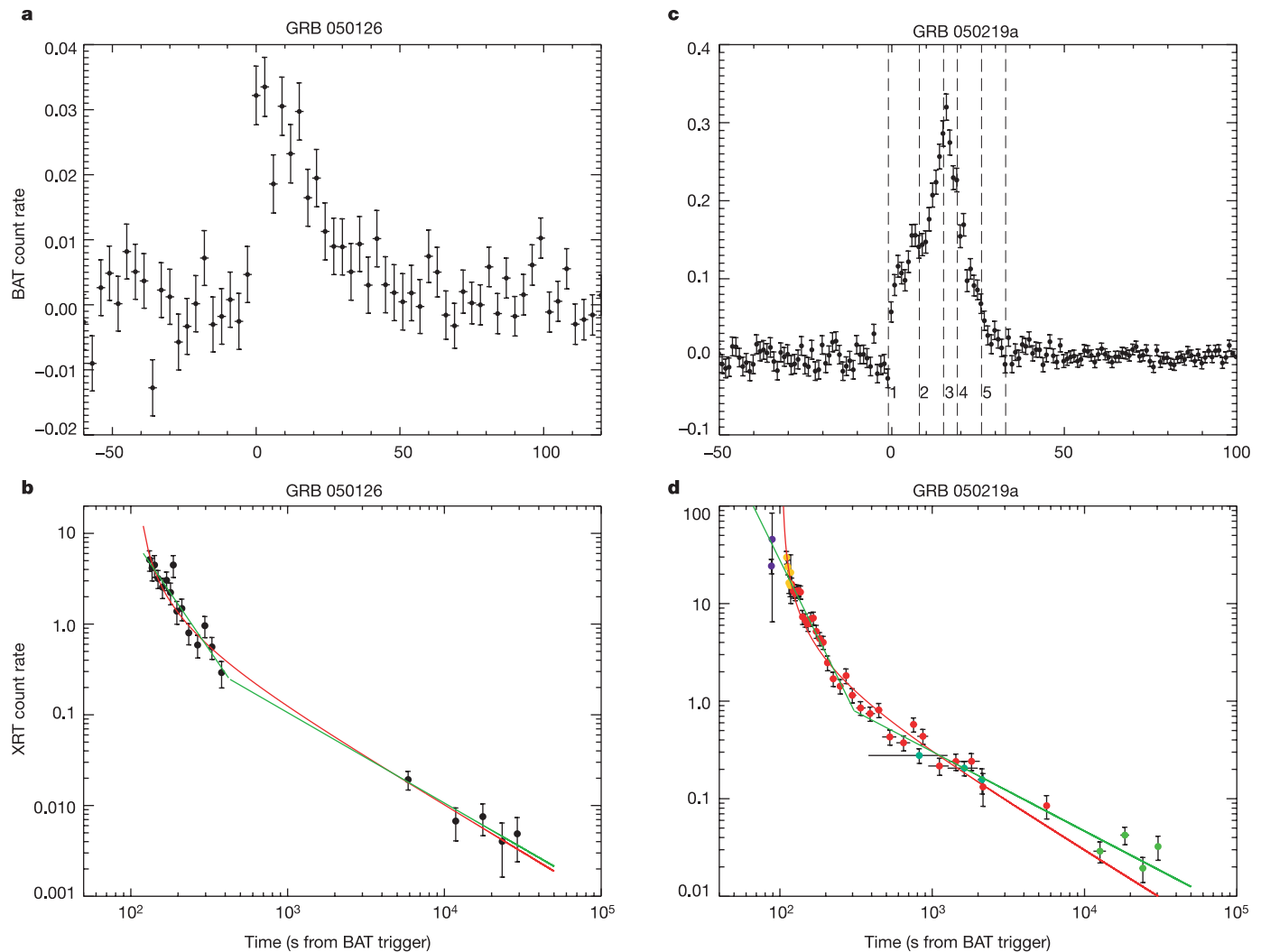


Figure 2 | The X-ray light curves of GRB050126 and GRB050219a as seen by BAT and XRT. Error bars represent the standard deviation. On January 26 2005, 12:00:54 UT, GRB050126 was detected and located by the BAT (a). Swift promptly slewed to the burst and settled at 129 s after the BAT trigger. Then the XRT took data until 12:07:42 UT, detecting a very bright and rapidly fading X-ray counterpart. GRB050126 was further observed by the XRT for the following eight orbits (b). On February 19 2005, 12:40:01 UT, GRB050219a triggered the BAT (c). Swift autonomously slewed to the BAT burst location and was on target after 87 s. The XRT executed the standard sequence of observations for GRBs³⁰, again detecting a very bright and rapidly fading X-ray counterparts across five orbits (d). In all plots, times are from the onset of the prompt emission. For GRB050219a this is 6 s earlier than the time reported in ref. 19. GRB050126 is a fast-rise-exponential-decay

and possibly three distinct mechanisms, as discussed below. For a self-similar forward shock solution in the standard GRB model, the time shift between the γ -ray trigger and the onset of afterglow emission is expected to be small. If this is within a few tens of seconds of the burst trigger time, which our data seem to indicate is true at least for GRB050219a, then we have a very steep initial decline of the early X-ray afterglow light curve, which requires explanation. A rapidly falling X-ray light curve at early times may arise in a hot cocoon accompanying a relativistic jet^{21,22}, or could be the photospheric emission associated with the outflow from the explosion²³. However, in the simplest versions of these models the spectrum of the emergent radiation is thermal, which is inconsistent with the power law spectrum observed for the two bursts. Some modifications to these models involving a comptonized power law tail of thermal radiation, for instance, might produce the observed behaviour. An

GRB with a total duration of ~ 30 s. GRB050219a has a more complex and multi-peaked light curve with a total duration of ~ 32 s. For this burst we performed the BAT spectral analysis over five intervals as shown. Panels b and d show the XRT light curves. The green lines represent a broken power law best fit to the afterglow decay, while the red lines represent the best fit for a model $\propto (t - t_0)^{-\alpha}$. For GRB050219a we also have an earlier detection before the decaying part (see blue points), which seems to indicate that the peak is in between these two points (not used in the fit) and the decaying part of the light curve, that is, in the range 88–110 s. The blue and yellow points are in photodiode mode, the red points are in window timing mode and the green points in photon counting mode (see ref. 30 for a description of the XRT operating modes).

alternative possibility is that the steep afterglow decay is produced in the external shock from a jet consisting of narrow regions of angular size $\leq \Gamma^{-1}$, where Γ is the jet Lorentz factor. As Γ decreases, the opening angle from which radiation can be seen becomes larger, without encompassing a larger fraction of the jet. A steep decay in the light curve is thus produced²⁴. Yet another possibility is light delay effects in off-axis emission ($\theta > \Gamma^{-1}$) from a relativistic jet arriving at the observer when emission from $\theta < \Gamma^{-1}$ has dropped to very small values owing to the adiabatic cooling of the shock heated shell²⁵, where θ is the angle between the line of sight to the observer and the normal to the jet surface.

A very interesting possibility is that the steep, early, X-ray light-curve is due to emission from the reverse shock heated ejecta^{25,26}. The peak of the synchrotron emission in the reverse shock is in the infrared or optical. These photons, if scattered by relativistic

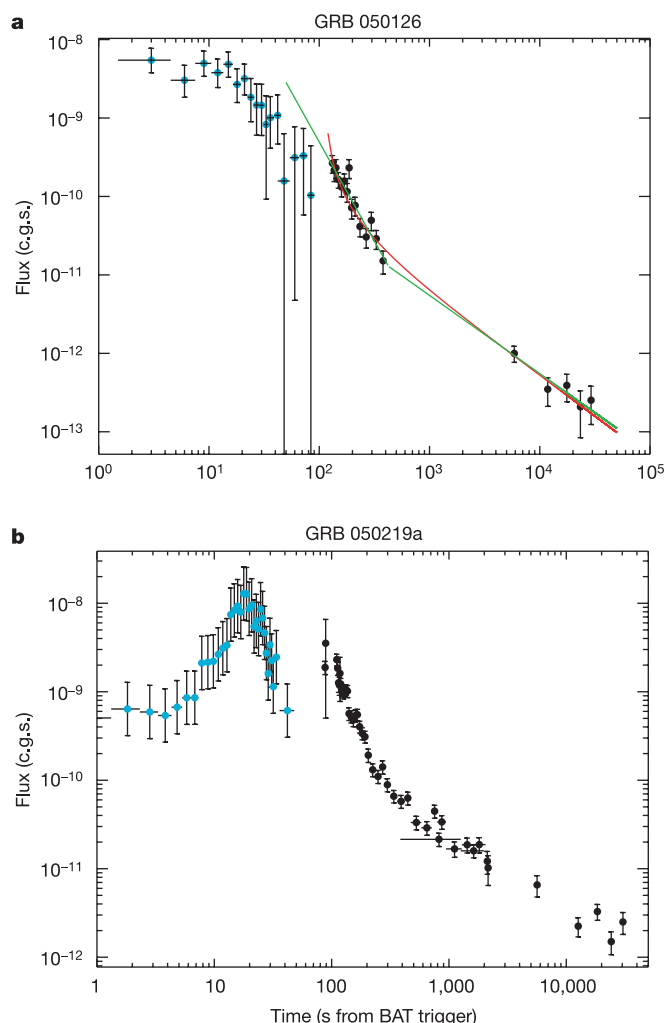


Figure 3 | Evolution of the two GRB X-ray light curves from the prompt phase to the afterglow phase. **a**, GRB 050126; **b**, GRB 050219a. The BAT and XRT count rates are converted into fluxes in a common energy band (0.2–10 keV). The conversion factors have been calculated using the best fit models that reproduced the BAT prompt spectra (for GRB050219a we used the values reported in Table 2) and the XRT afterglow spectra, respectively. The error bars represent the standard deviation plus the estimated uncertainties in the conversion factors. For GRB050219a the X-ray source detected by XRT is at a higher level than the late stages of the prompt emission: there is a clear discontinuity between the BAT and XRT light curves. The XRT light curve also shows a hint of a rising phase before the onset of the decay. Note that the BAT detector is taking data all the time. We stop plotting them after ~ 80 and 50 s, for the two GRBs respectively, because the sources are not detected any more. For a 5σ detection of the GRB050219a X-ray source seen by XRT at ~ 90 – 100 s, BAT would need more than 100 s. Given that this source is rapidly fading, it is too weak to be detected by BAT. For GRB050126 the X-ray flux is weaker and the discontinuity is not so evident. Moreover, the BAT conversion factor is calculated over the averaged spectrum. For a strong spectral evolution from hard to soft the latter BAT points would have higher fluxes. However, the 5σ peak spectrum and the averaged total spectrum have very similar spectral indexes¹⁸, so we do not have indication of a strong spectral evolution. In conclusion, for GRB050126 the BAT and XRT light curves do not seem to simply connect as well, although for this GRB this cannot be ruled out.

electrons in the ejecta, emerge in the X-ray band. The X-ray light curve in this case will decline roughly as $t^{-2.6}$, which is consistent with observations. However, in order to avoid very bright early optical radiation from these bursts, which was not seen, the ejecta may need to be highly enriched with electron/positron pairs, with an ejecta Lorentz factor of at least a few hundred.

We note that none of these models are completely consistent with all the available data in the γ -rays, X-rays, and optical upper limits for these two GRBs, which suggests the need for refining the current models.

Received 22 April; accepted 14 June 2005.

1. Mészáros, P. & Rees, M. Optical and long-wavelength afterglow from gamma-ray bursts. *Astrophys. J.* **476**, 232–237 (1999).
2. Sari, R. & Piran, T. GRB 990123: The optical flash and the fireball model. *Astrophys. J.* **517**, L109–L112 (1999).
3. Ford, L. A. *et al.* BATSE observations of gamma-ray burst spectra. 2: Peak energy evolution in bright, long bursts. *Astrophys. J.* **439**, 307–321 (1995).
4. Frontera, F. *et al.* Spectral properties of the prompt X-ray emission and afterglow from the gamma-ray burst of 1997 February 28. *Astrophys. J.* **439**, L67–L70 (1998).
5. Zhang, B. & Mészáros, P. Gamma-ray bursts: progress, problems, and prospects. *Int. J. Mod. Phys.* **19**, 2385–2472 (2004).
6. Piran, T. The physics of gamma-ray bursts. *Rev. Mod. Phys.* **76**, 1143–1210 (2005).
7. Piro, L. *et al.* Evidence for a late-time outburst of the X-ray afterglow of GRB970508 from BeppoSAX. *Astron. Astrophys.* **331**, L41–L44 (1998).
8. Giblin, T. W. *et al.* Evidence for an early high-energy afterglow observed with BATSE from GRB 980923. *Astrophys. J.* **524**, L47–L50 (1999).
9. Burenin, R. A. *et al.* GRANAT/SIGMA observation of GRB 920723 soft gamma-ray afterglow. *Astron. Astrophys.* **344**, L53–L56 (1999).
10. Piro, L. *et al.* Probing the environment in gamma-ray bursts: The case of an X-ray precursor, afterglow late onset and wind versus constant density profile in GRB011121 and GRB011211. *Astrophys. J.* **623**, 314–324 (2005).
11. Akerlof, C. *et al.* Observations of contemporaneous optical radiation from a gamma-ray burst. *Nature* **398**, 400–402 (1999).
12. Fox, D. W. *et al.* Discovery of early optical emission from GRB021211. *Astrophys. J.* **586**, L5–L8 (2003).
13. Gehrels, N. *et al.* The *Swift* gamma-ray burst mission. *Astrophys. J.* **611**, 1005–1020 (2004).
14. Burrows, D. N. *et al.* *Swift* X-ray telescope and Very Large Telescope observations of the afterglow of GRB 041223. *Astrophys. J.* **622**, L85–L88 (2005).
15. Campana, S. *et al.* *Swift* observations of GRB050128: the early X-ray afterglow. *Astrophys. J.* **625**, L23–L26 (2005).
16. Burrows, D. N. *et al.* The *Swift* X-ray telescope. *Space Sci. Rev.* (in the press).
17. Barthelmy, S. D. *et al.* The Burst Alert Telescope (BAT) on the *Swift* MIDEX mission. *Space Sci. Rev.* (in the press).
18. Sato, G. *et al.* GRB050126: Bursts parameters measured by BAT. *GCN Circ.* 2987 (2005).
19. Hullinger, D. *et al.* GRB050219: *Swift*-BAT detection of a burst. *GCN Circ.* 3038 (2005).
20. McMahon, E., Kumar, P. & Panaitescu, A. Prompt gamma-ray and early afterglow emission in the external shock model. *Mon. Not. R. Astron. Soc.* **354**, 915–923 (2004).
21. Mészáros, P. & Rees, M. J. Collapsar jets, bubbles, and Fe lines. *Astrophys. J.* **556**, L37–L40 (2001).
22. Ramirez-Ruiz, E., Celotti, A. & Rees, M. J. Events in the life of a cocoon surrounding a light, collapsar jet. *Mon. Not. R. Astron. Soc.* **337**, 1349–1356 (2002).
23. Rees, M. J. & Mészáros, P. Dissipative photosphere models of gamma-ray bursts and X-ray flashes. *Astrophys. J.* (in the press); (preprint at <http://arXiv.org/astro-ph/0412702>) (2005).
24. Kumar, P. & Piran, T. Energetics and luminosity function of gamma-ray bursts. *Astrophys. J.* **535**, 152–157 (2000).
25. Kumar, P. & Panaitescu, A. Afterglow emission from naked gamma-ray bursts. *Astrophys. J.* **541**, L51–L54 (2000).
26. Vietri, M. A simple, stringent test on the nature of GRB progenitors. *Astrophys. J.* **14**, 211–215 (2000).
27. Roming, P. W. *et al.* The *Swift* ultra-violet/optical telescope. *Space Sci. Rev.* (in the press).
28. Berger, E. *et al.* The discovery of the optical and near-IR afterglows of the first *Swift* gamma-ray bursts. *Astrophys. J.* (in the press); (preprint at <http://arXiv.org/astro-ph/0502468>) (2005).
29. Schady, P. *et al.* GRB050219a: No *Swift* UVOT detection of afterglow emission. *GCN Circ.* 3039 (2005).
30. Hill, J. E. *et al.* Read-out modes and automated operation of the *Swift* X-ray telescope. *Proc. SPIE* **5165**, 217–231 (2004).

Acknowledgements We acknowledge support from ASI, NASA and PPARC.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.T. (tagliaferri@merate.mi.astro.it).

Young chondrules in CB chondrites from a giant impact in the early Solar System

Alexander N. Krot¹, Yuri Amelin², Patrick Cassen³ & Anders Meibom⁴

Chondrules, which are the major constituent of chondritic meteorites, are believed to have formed during brief, localized, repetitive melting of dust (probably caused by shock waves^{1,2}) in the protoplanetary disk around the early Sun. The ages of primitive chondrules^{3–6} in chondritic meteorites indicate that their formation started shortly after that of the calcium–aluminium-rich inclusions (4,567.2 ± 0.7 Myr ago) and lasted for about 3 Myr, which is consistent with the dissipation timescale for protoplanetary disks around young solar-mass stars⁷. Here we report the ²⁰⁷Pb–²⁰⁶Pb ages of chondrules in the metal-rich CB (Bencubbin-like) carbonaceous chondrites Gujba (4,562.7 ± 0.5 Myr) and Hammadah al Hamra 237 (4,562.8 ± 0.9 Myr), which formed during a single-stage, highly energetic event^{8–11}. Both the relatively young ages and the single-stage formation of the CB chondrules are inconsistent with formation during a nebular shock wave². We conclude that chondrules and metal grains in the CB chondrites formed from a vapour–melt plume produced by a giant impact between planetary embryos after dust in the protoplanetary disk had largely dissipated. These findings therefore provide evidence for planet-sized objects in the earliest asteroid belt, as required by current numerical simulations of planet formation in the inner Solar System¹².

Chondrules are igneous, spherical objects, ~0.01–10 mm in size, composed largely of ferromagnesian olivine and pyroxene, metal (Fe, Ni), and glassy or fine-grained silicate material (mesostasis). Most chondrules have porphyritic textures, which are consistent with crystal growth from a rapidly cooling (100–1,000 K h^{−1}) silicate melt^{2,13}. Chondrules are often surrounded by coarse-grained igneous rims or contain relict fragments of earlier generations of chondrules, indicating that chondrule formation was a repetitive process^{1,2,13}. Additionally, chondrules in most chondrites are surrounded by fine-grained, matrix-like rims^{2,13}. On the basis of these observations, it is generally inferred that chondrules are quenched droplets of a liquid that formed by varying degrees of melting of dense aggregates of ferromagnesian silicate, metal and sulphide grains during repetitive flash-heating events in the dusty solar nebula^{1,2,13}. Nebular shock waves are currently the favourite mechanism for chondrule formation²; however, the sources of shock waves in the solar nebula have yet to be identified¹⁴. The proposed sources include accretion shock¹⁵, infalling clumps of dust and gas¹⁶, bow shocks generated by planetesimals¹⁷, spiral arms and clumps in a gravitationally unstable protoplanetary disk^{14,18}, and X-ray flares¹⁹. Here we report the mineralogy, petrography and ²⁰⁷Pb–²⁰⁶Pb ages of chondrules in the CB carbonaceous chondrites Gujba and Hammadah al Hamra 237 (HH 237 hereafter), and show that they are inconsistent with an origin in shock wave heating in the solar nebula and instead require a fundamentally different formation mechanism.

The CB chondrites comprise a diverse group of five meteorites (Bencubbin, Gujba, Weatherford (BGW hereafter), HH 237 and Queen Alexandra Range (QUE) 94411), which are characterized by similar oxygen isotopic compositions, high (60–70 vol.%) abundance of Fe, Ni-metal ± sulphide, extreme depletion in moderately volatile elements, and extreme enrichment in δ¹⁵N (ref. 20). HH 237 and QUE 94411 are finer-grained than BGW, and contain rare, uniformly ¹⁶O-poor refractory inclusions and abundant chemically zoned Fe, Ni-metal grains, which are absent in BGW (Fig. 1). The compositional zoning of the metal grains is consistent with a gas–solid condensation origin^{21–23}. In contrast to typical chondrules in ordinary, enstatite and carbonaceous chondrites^{1,13}, chondrules in CB chondrites have exclusively non-porphyritic (skeletal olivine and cryptocrystalline) textures and magnesium-rich compositions, and lack relict grains, coarse-grained igneous rims, and fine-grained matrix-like rims^{8–10}. They are highly depleted in moderately volatile elements (Mn, Na, K) and have unfractionated refractory lithophile (Ca, Al, Ti and rare earth elements) abundance patterns^{8,10}. Chondrules in HH 237 exhibit a large range (4 to <0.01 times CI carbonaceous chondrite abundances) in bulk abundances of refractory lithophile elements among individual chondrules compared to those in Gujba (1.0–1.6 × CI) (refs 8, 10). The CB chondrules are metal-free, but small cryptocrystalline chondrules are commonly observed inside chemically zoned metal condensates in HH 237 and QUE 94411 (ref. 8). Although there are significant variations in grain sizes among Gujba chondrules as well, no strictly cryptocrystalline chondrules were found; all chondrules have skeletal olivine textures (see Supplementary Fig. 2).

The origin of CB meteorites remains highly controversial: nebular^{8,21,22} and asteroidal^{9,24} models have been proposed. Some authors^{8,21,22} reached the conclusion that chondrules and zoned metal grains in HH 237 and QUE 94411 record a highly energetic thermal event resulting in nearly complete evaporation of a dusty region of the solar nebula, after which chondrules condensed as liquid droplets, followed by gas–solid condensation of zoned Fe, Ni-metal grains. Because the origin of chondrules and metal grains in HH 237 and QUE 94411 requires a very energetic astrophysical environment, these authors postulated that they formed earlier than chondrules and metal in other chondrites, when the protoplanetary disk in the inner Solar System was dense and hot as a result of high mass-accretion rates to the Sun. Although condensation (or evaporation) has also been invoked for the origin of the larger, homogeneous metal ± sulphide nodules in BGW^{9,24} (Fig. 1; see also Supplementary Fig. 1), this may require a gas with extremely high partial pressures of the siderophile elements (up to ~10⁷ × CI). Rubin *et al.*⁹ and Campbell *et al.*²⁴ hypothesized that metal ± sulphide and silicate nodules in BGW formed by gas–liquid condensation or evaporation

¹Hawai'i Institute of Geophysics & Planetology, School of Ocean & Earth Science & Technology, University of Hawai'i at Manoa, 2525 Correa Rd, Honolulu, Hawaii 96822, USA.

²Geology Department, University of Toronto, and Geological Survey of Canada, 601 Booth St, Ottawa, Ontario, K1A 0E8, Canada. ³SETI Institute, 515 N. Whisman Road, Mountain View, California 94043, USA. ⁴Museum National d'Histoire Naturelle, Laboratoire d'Etude de la Matière Extraterrestre, USM 0205, Case Postale 52, 57 rue Cuvier, 75005 Paris, France.

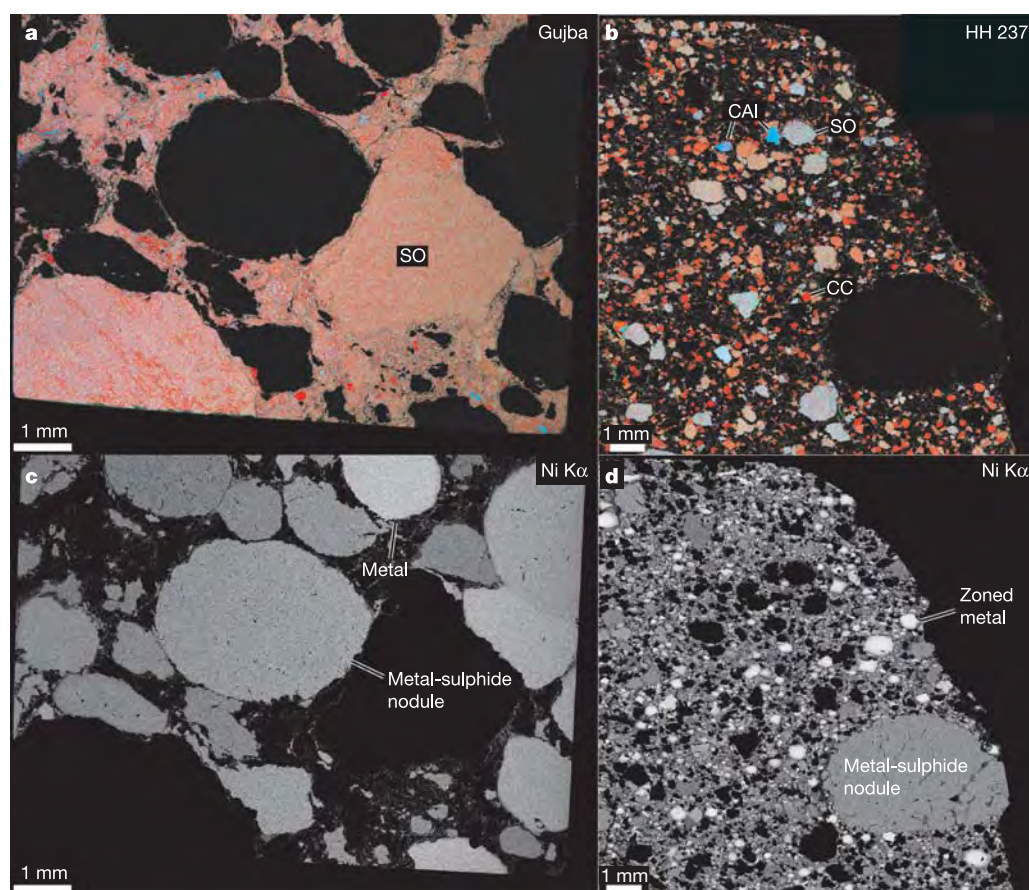


Figure 1 | The CB carbonaceous chondrites Hammadah al Hamra 237 (HH 237) and Gujba. **a, b**, Combined elemental maps in Mg (red), Ca (green) and Al K α (blue) X-rays. **c, d**, Elemental maps in Ni K α X-rays. **a, c**, Gujba consists of large chondrule fragments with skeletal olivine (SO) textures and Fe,Ni $\pm$ sulphide nodules; metal nodules are compositionally uniform. **b, d**, HH 237 contains abundant, small Fe,Ni-metal grains,

chondrules, and calcium-aluminium-rich inclusions (CAIs), and rare, large Fe,Ni $\pm$ sulphide nodules. Chondrules have either cryptocrystalline (CC, reddish colours) or skeletal olivine (SO, bluish colours) textures. Rare, large SO chondrules were found in a polished slab of HH 237 (see Supplementary Fig. 1).

in a high-density, metal-enriched vapour cloud generated in a protoplanetary impact.

Five chondrules with skeletal olivine textures were extracted from Gujba and characterized by scanning electron microscopy and electron probe microanalysis. These chondrules were subsequently crushed and analysed for Pb concentrations and Pb isotopic compositions. Analytical procedures and results are presented in Supplementary Information. Fragments from three chondrules have highly radiogenic Pb isotopic compositions with measured $^{206}\text{Pb}/^{204}\text{Pb}$ ratios of 307–2,602 (four out of 11 analysed fractions yielded $^{206}\text{Pb}/^{204}\text{Pb} > 1,000$) and tightly correlated Pb isotopic ratios. Common Pb content in most of these fractions is close to the analytical blank, and the influence of common Pb heterogeneity on the age calculation is therefore very small. Isochron regression of these data (Fig. 2) yielded an age of $4,562.7 \pm 0.5$ Myr (mean square of weighted deviates, MSWD = 1.3), which we consider the best estimate for the timing of formation of the Gujba silicate chondrules. Consistency of radiogenic $^{207}\text{Pb}/^{206}\text{Pb}$ between analysed fractions strongly suggests single-stage evolution without noticeable loss of radiogenic Pb during secondary processing. There are many factors (for example, grain size, mineral composition, fracturing) that can make shock- or alteration-related Pb loss vary, but there is no process known to stabilize the degree of secondary Pb loss.

Seven fractions of chondrule fragments and two individual chondrules separated from HH 237, as well as acid leachates of these fractions, have been analysed for U and Pb concentrations and Pb isotopic compositions. Analytical procedures and results

are presented in Supplementary Information. Five fractions separated by dissolution of metal with subsequent removal of remaining magnetic phases yielded a Pb–Pb isochron age of $4,564.3 \pm 2.0$ Myr (MSWD = 1.19). Two-point isochron dates for ‘residue-second acid wash’ pairs (see Supplementary Information for details) for these fractions are identical within error. The weighted average of these dates of $4,562.8 \pm 0.9$ Myr (MSWD = 0.44) is consistent with the isochron date but is more precise, and represents our current best estimate for the timing of formation of the HH 237 silicates. Thus the CB chondrules postdate formation of calcium-aluminium-rich inclusions (CAIs) in CV chondrites by about 5 Myr (ref. 3).

Our results establish contemporaneous origin and strongly suggest co-genesis of HH 237 and Gujba. The ^{207}Pb – ^{206}Pb ages of the CB chondrules are similar to the ^{182}Hf – ^{182}W ages of the metal–silicate fractionation recorded by the CB metal condensates²⁵, further substantiating a co-genetic origin for chondrules and metal in the CB meteorites. The very young (in comparison with CAIs) ages of the CB chondrules and metal grains produced by an extremely energetic, single-stage formation mechanism, and the lack of inter-chondrule fine-grained matrix material (nebular dust; see Supplementary Fig. 3) are inconsistent with the nebular shock wave heating mechanisms currently favoured for the origin of typical chondrules^{2,14}. The latter mechanisms are repeatable processes that result in incomplete melting and evaporation of chondrule precursors, and evaporation and re-condensation of fine-grained matrix materials². The lack of inter-chondrule fine-grained matrix materials in CB chondrites may indicate that the fine-grained nebular dust had largely agglomerated

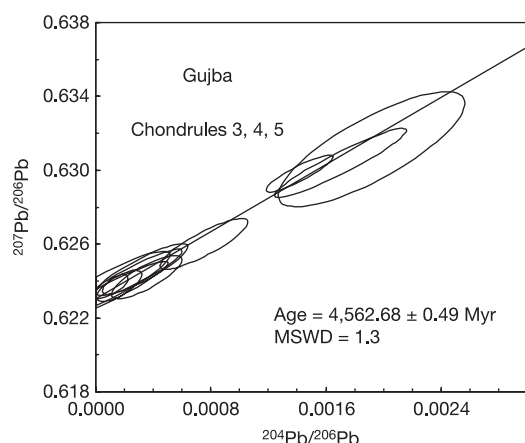


Figure 2 | $^{204}\text{Pb}/^{206}\text{Pb}$ – $^{207}\text{Pb}/^{206}\text{Pb}$ isochron diagram for three Gujba chondrules (numbers 3, 4 and 5). $^{207}\text{Pb}/^{206}\text{Pb}$ ratios are not corrected for initial common Pb. Error ellipses are 2σ . Isochron age errors are 95% confidence intervals. MSWD, mean square of weighted deviates.

into planetesimals by the time of the CB chondrite formation, consistent with the disappearance of infrared excesses in the spectral energy distributions of young solar-mass stars over a timescale of 1–3 Myr (ref. 7).

We conclude that the young formation ages of the CB chondrites and their anomalous mineralogical and chemical characteristics discussed above are inconsistent with a nebular setting^{8,11,21,22} but strongly support an origin of these meteorites in a big protoplanetary impact^{9,10,24}. In such an impact, melting results from shock compression, which is followed by adiabatic decompression; expansion causes the supercritical fluid to boil, producing melt droplets that are accelerated by the expanding gases²⁶. We suggest that large skeletal olivine chondrules and metal nodules in Gujba and HH 237 are melts produced by such a collision^{9,24}, whereas the smaller cryptocrystalline chondrules and chemically zoned metal grains in HH 237 are respectively gas–liquid and gas–solid condensates from a vapour phase^{8,11,21–23}. Unmelted fragments of the CB chondrite precursor materials have not been observed in Gujba and HH 237 (refs 8, 9, 20). Melosh *et al.*²⁶ estimated that a collision between Moon-sized ($\sim 1,700$ km in radius) bodies would produce abundant vapour and melt, which could be efficiently separated from the fragmented rocks under the gravity field of the colliding bodies.

Although no such bodies currently exist in the asteroid belt (the largest asteroid is Ceres, ~ 930 km in diameter), models suggest that the present asteroid belt contains $<0.1\%$ of its original mass, most of which was stored in the form of Moon- to Mars-sized objects, which were subsequently removed through gravitational perturbations, close encounters and planetary resonances¹². Such embryos could have formed in the asteroid belt in substantially less than a million years²⁷. Because their removal was driven primarily by resonant interactions with Jupiter and Saturn, the timing and rate of removal depended on the formation and orbital eccentricities of these planets. Complete removal of large planetary bodies from the asteroid belt might have taken as long as a few hundred Myr (ref. 12), but their numbers would have diminished by half within only one or two Myr after the formation of Jupiter²⁸. Thus we expect that energetic collisions among them would have been most frequent soon after the formation of the gas giant planets, which must have occurred before the disappearance of nebular gas. It is therefore possible that some nebular gas and dust were still present when the CB meteorites formed. In fact, it must be considered likely that giant impacts occurred in the asteroid belt even before nebular dust was removed; perhaps a record of such events exists but remains unrecognized in older meteorites.

A giant impact with a Mars-sized protoplanet is commonly

inferred to explain the origin of the Earth's Moon^{29,30}. Although the Moon-forming event occurred much later (42 ± 4 Myr later; ref. 29), and may have been more energetic than that envisioned here, simulations of it yield some relevant conclusions³⁰. First, the products of such a collision are typically a massive primary surrounded by orbiting material, plus a modest fraction of material that escapes the local system, and so does not immediately become part of the merged protoplanet. Second, as mentioned above, abundant vapour and melt are produced. Third, vapour and melt tend to dominate the escaping material. The ultimate survival of the escaping material is favoured over that of the massive primary to the extent that it eventually coalesces to form (or is accreted by) numerous small objects. That is, since ejection from the belt is size-independent, most of the surviving objects were the more numerous smaller bodies, while most of the mass was lost in ejected large objects that dominated the mass distribution.

The high-precision absolute ages of the CB chondrules (and metal) and their unique single-stage formation mechanism have important implications for isotopic dating. This formation event has probably homogenized radionuclides in chondrules and metal of the CB chondrites, and reset short-lived radiogenic isotope systems, such as ^{26}Al – ^{26}Mg (half-life, $t_{1/2} = 0.74$ Myr), ^{60}Fe – ^{60}Ni ($t_{1/2} = 1.5$ Myr), ^{53}Mn – ^{53}Cr ($t_{1/2} = 3.7$ Myr), ^{107}Pd – ^{107}Ag ($t_{1/2} = 6.5$ Myr) and ^{182}Hf – ^{182}W ($t_{1/2} = 9$ Myr). These isotope systems can be used to determine the relative ages of the earliest events in the Solar System (for example, chondrule and CAI formation, thermal metamorphism, aqueous alteration, igneous differentiation and core formation in asteroids and planets)^{4–6}. For establishing consistent Solar System chronology, these chronometers have to be linked together and tied to an absolute timescale^{5,6}. Most meteorites are made of components formed at different time, and/or experienced complex and prolonged post-formation metamorphic history, and are not suitable for linking short-lived chronometers. In contrast, the correlated studies of multiple short-lived isotope systems in CB chondrites can potentially test the consistency among them and provide a tie to an absolute timescale, which will be an important step towards the unified timescale of the earliest Solar System.

Received 3 April; accepted 20 May 2005.

1. Scott, E. R. D. & Krot, A. N. in *Meteorites, Comets, and Planets* (ed. Davis, A. M.) Vol. 1 *Treatise on Geochemistry* (eds Holland, H. D. & Turekian, K. K.) 327–334 (Cambridge Univ. Press, Cambridge, 2003).
2. Desch, S. J. & Connolly, H. C. Jr. A model of the thermal processing of particles in solar nebula shocks: Application to the cooling rates of chondrules. *Meteorit. Planet. Sci.* **37**, 183–207 (2002).
3. Amelin, Y., Krot, A. N., Hutcheon, I. D. & Ulyanov, A. A. Lead isotopic ages of chondrules and calcium–aluminum-rich inclusions. *Science* **297**, 1678–1683 (2002).
4. Bizzarro, M., Baker, J. A. & Haack, H. Mg isotope evidence for contemporaneous formation of chondrules and refractory inclusions. *Nature* **431**, 275–278 (2004).
5. McKeegan, K. D. & Davis, A. M. in *Meteorites, Comets, and Planets* (ed. Davis, A. M.) Vol. 1 *Treatise on Geochemistry* (eds Holland, H. D. & Turekian, K. K.) 431–461 (Cambridge Univ. Press, Cambridge, 2003).
6. Kita, N., *et al.* in *Chondrites and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) (Astrophysical Society of the Pacific, in the press).
7. Hartmann, L. in *Chondrites and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) (Astrophysical Society of the Pacific, in the press).
8. Krot, A. N. *et al.* A new astrophysical setting for chondrule formation. *Science* **291**, 1776–1779 (2001).
9. Rubin, A. E. *et al.* Formation of metal and silicate globules in Gujba: A new Bencubbin-like meteorite fall. *Geochim. Cosmochim. Acta* **67**, 3283–3298 (2003).
10. Krot, A. N., Amelin, Y., Russell, S. S. & Twelker, E. Are chondrules in the CB carbonaceous chondrite Gujba primary (nebular) or secondary (asteroidal)? *Meteorit. Planet. Sci.* **39**, A56 (2004).
11. Meibom, A. J. *et al.* Large scale thermal events in the solar nebula recorded in Fe,Ni metal condensates in primitive meteorites. *Science* **288**, 839–841 (2000).
12. Chambers, J. E. & Wetherill, G. W. Planets in the asteroid belt. *Meteorit. Planet. Sci.* **36**, 381–399 (2001).
13. Jones, R. H., Lee, T., Connolly, H. C. Jr, Love, S. G. & Shang, H. in *Protostars and Planets IV* (eds Manning, V., Boss, A. P. & Russell, S. S.) 927–962 (Univ. Arizona Press, Tucson, 2000).

14. Boss, A. P. & Durisen, R. H. in *Chondrites and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) (Astrophysical Society of the Pacific, in the press).
15. Wood, J. A. On the formation of meteoritic chondrules by aerodynamic drag heating in the solar nebula. *Earth Planet. Sci. Lett.* **70**, 11–26 (1984).
16. Boss, A. P. & Graham, J. A. Clumpy disk accretion and chondrule formation. *Icarus* **106**, 168–178 (1993).
17. Weidenschilling, S. J., Marzari, F. & Hood, L. L. The origin of chondrules and jovian resonances. *Science* **279**, 681–684 (1998).
18. Wood, J. A. Processing of chondritic and planetary materials in spiral density waves in the nebula. *Meteorit. Planet. Sci.* **31**, 641–645 (1996).
19. Nakamoto, T., Hayashi, M. R., Kita, N. T. & Tachibana, S. in *Chondrites and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) (Astrophysical Society of the Pacific, in the press).
20. Weisberg, M. K. et al. A new metal-rich chondrite group. *Meteorit. Planet. Sci.* **36**, 401–418 (2001).
21. Petaev, M. I., Meibom, A., Krot, A. N., Wood, J. A. & Keil, K. The condensation origin of zoned metal grains in Queen Alexandra Range 94411: Implications for the formation of the Bencubbin-like chondrites. *Meteorit. Planet. Sci.* **36**, 93–106 (2001).
22. Campbell, A. J., Humayun, M., Meibom, A. J., Krot, A. N. & Keil, K. Origin of zoned metal grains in the QUE94411 chondrite. *Geochim. Cosmochim. Acta* **65**, 163–180 (2001).
23. Alexander, C. M. O'D. & Hewins, R. H. Mass fractionation of Fe and Ni isotopes in metal in Hammadah al Hamra 237. *Meteorit. Planet. Sci.* **39**, A13 (2004).
24. Campbell, A. J., Humayun, M. & Weisberg, M. K. Siderophile element constraints on the formation of metal in the metal-rich chondrites Bencubbin, Weatherford, and Gujba. *Geochim. Cosmochim. Acta* **66**, 647–660 (2002).
25. Kleine, T., Mezger, K., Palme, H., Scherer, E. & Münker, C. Early core formation in asteroids and late accretion of chondrite parent bodies: Evidence from ^{182}Hf – ^{182}W in CAIs, metal-rich chondrites and iron meteorites. *Geochim. Cosmochim. Acta* (submitted).
26. Melosh, H. J., Cassen, P., Sears, D. & Lugmair, G. in *Workshop on Chondrites and the Protoplanetary Disk* (eds Krot, A. N., Reipurth, B. & Scott, E. R. D.) 125–126 (Hawaii Institute of Geophysics and Planetology, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, 2004).
27. Wetherill, G. W. & Stewart, G. Formation of planetary embryos: effects of fragmentation, low relative velocity, and independent variation of eccentricity and inclination. *Icarus* **106**, 190–209 (1993).
28. Chambers, J. E. & Cassen, P. The effects of nebula surface density profile and giant-planet eccentricities on planetary accretion in the inner solar system. *Meteorit. Planet. Sci.* **37**, 1523–1540 (2002).
29. Kleine, T., Mezger, K. & Palme, H. The Hf–W age of the lunar magma ocean. *Lunar Planet. Sci.* XXXVI, 1940 (2005).
30. Canup, R. M. Simulations of a late lunar-forming impact. *Icarus* **168**, 433–456 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support for this project was provided by NASA (A.N.K., P.C.) and the Canadian Space Agency (Y.A.). We thank J. A. Wood for comments and suggestions, and E. Twelker for providing samples and images of Gujba and Hammadah al Hamra 237.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.N.K. (sasha@higp.hawaii.edu).

Experimental measurement of the photonic properties of icosahedral quasicrystals

Weining Man^{1,2}, Mischa Megens³, Paul J. Steinhardt¹ & P. M. Chaikin^{1,2,4}

Quasicrystalline structures may have optical bandgap properties—frequency ranges in which the propagation of light is forbidden—that make them well-suited to the scientific and technological applications for which photonic crystals^{1–3} are normally considered⁴. Such quasicrystals can be constructed from two or more types of dielectric material arranged in a quasiperiodic pattern whose rotational symmetry is forbidden for periodic crystals (such as five-fold symmetry in the plane and icosahedral symmetry in three dimensions). Because quasicrystals have higher point group symmetry than ordinary crystals, their gap centre frequencies are closer and the gaps widths are more uniform—optimal conditions for forming a complete bandgap that is more closely spherically symmetric. Although previous studies have focused on one-dimensional and two-dimensional quasicrystals^{4–7}, where exact (one-dimensional) or approximate (two-dimensional) band structures can be calculated numerically, analogous calculations for the three-dimensional case are computationally challenging and have not yet been performed. Here we circumvent the computational problem by doing an experiment. Using stereolithography, we construct a photonic quasicrystal with centimetre-scale cells and perform microwave transmission measurements. We show that three-dimensional icosahedral quasicrystals exhibit sizeable stop gaps and, despite their quasiperiodicity, yield uncomplicated spectra that allow us to experimentally determine the faces of their effective Brillouin zones. Our studies confirm that they are excellent candidates for photonic bandgap materials.

In 1984, Schechtman *et al.* observed icosahedral symmetry with five-fold rotation axes in the electron diffraction pattern of an alloy of Al-Mn (ref. 8). Simultaneously, the concept of long range quasiperiodic order with icosahedral symmetry was theoretically developed by Levine and Steinhardt^{9,10}. Our realization of a photonic icosahedral quasicrystal is shown in Fig. 1a. The diamond structure in Fig. 1b was made for comparative experiments; diamond has been suggested as an optimal structure for photonic crystals.

Photonic crystals are based on the fact that photons Bragg-scatter from a medium with a periodically modulated refractive index. Multiple scattering at frequencies near the Bragg condition prevents propagation in these directions, producing a ‘stop gap’. Overlap of the stop gaps in all directions yields a complete photonic bandgap and traps the light. Intuitively, the complete overlap occurs more readily in more isotropic structures. Quasicrystals have long-range quasiperiodic order and higher point group symmetries, so photons Bragg-scatter along a more spherically symmetric set of directions. Many recent papers address this question in two dimensions^{4,6,7}. As the symmetry increases, the Brillouin zone becomes more circular or more spherical. Photonic quasicrystals also allow for a higher degree of flexibility and tunability for defect mode properties⁶.

Figure 1c shows the effective Brillouin zone (related to the pseudo-Jones zone used in describing electronic transport in quasicrystals^{11,12}) of the icosahedral structure with its irreducible Brillouin zone highlighted in yellow. For comparison, Fig. 1d shows the first Brillouin zone of the diamond (face-centred cubic, f.c.c.) structure with its irreducible Brillouin zone. Note that, as a measure of sphericity, along the edge of the diamond structure’s irreducible Brillouin zone the magnitude of $\mathbf{k}$ (which is proportional to the stop gap centre frequency to a first-order approximation) increases 29.1% from L to W. Along the edge of the effective irreducible triacontahedral Brillouin zone of the icosahedral structure, the magnitude of $\mathbf{k}$ increases only 17.5% from the two-fold to the five-fold symmetry points. Moreover the triacontahedron’s faces are identical and subtend smaller solid angles.

A D -dimensional periodic lattice has D independent basis vectors, whereas $D + N$ linearly independent vectors (with integer $N \geq 1$

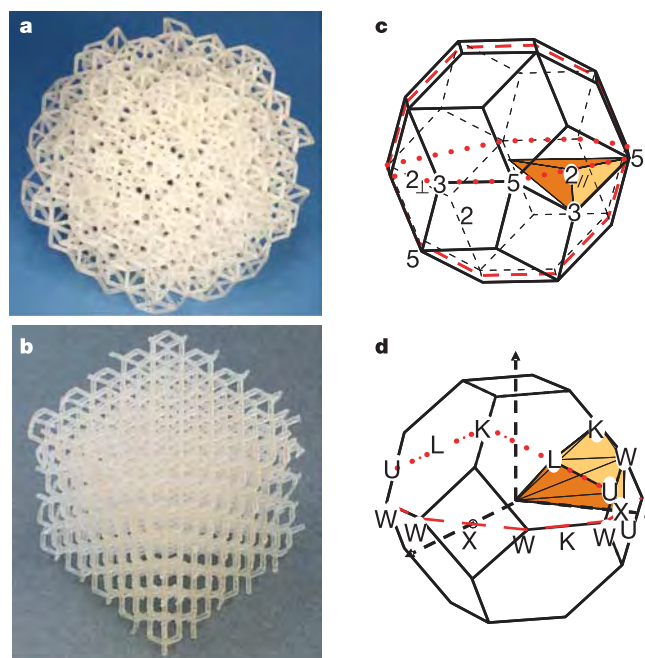


Figure 1 | Experimental photonic structures and their Brillouin zones. **a**, Stereolithographically produced icosahedral quasicrystal with 1-cm-long rods. **b**, Diamond structure with 1-cm-long rods. **c**, Triacontahedron, one of several possible effective Brillouin zones with icosahedral symmetry. **d**, Brillouin zone for the f.c.c./diamond structure.

¹Department of Physics, ²Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, New Jersey, 08544, USA. ³Philips Research Laboratories, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands. ⁴Department of Physics and Center for Soft Condensed Matter Research, New York University, New York 10003, USA.

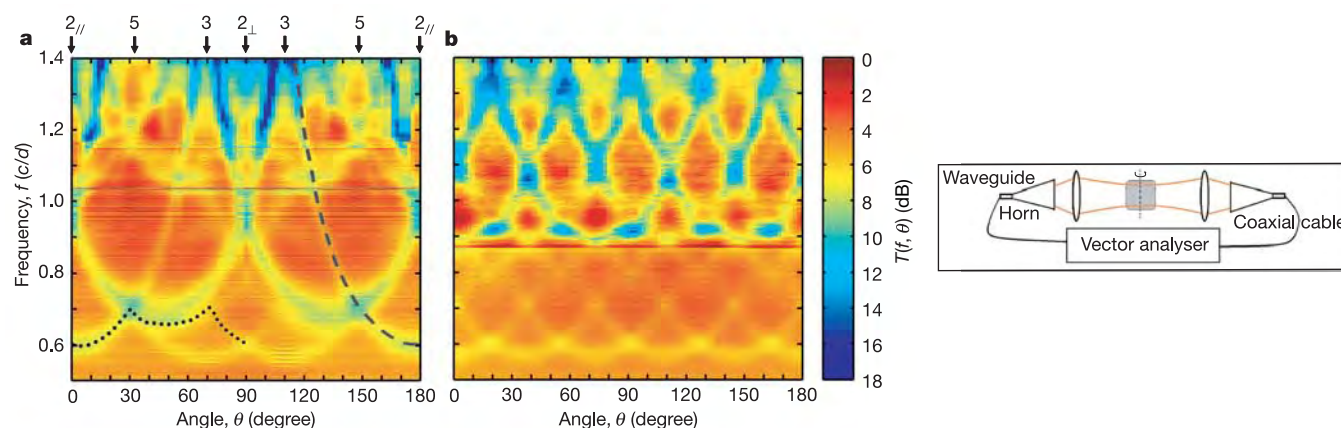


Figure 2 | Measured transmission for an icosahedral quasicrystal. a, $T(f, \theta)$, transmission as a function of frequency (measured in units of c/d) and angle, for a rotation about a two-fold rotation axis of the quasicrystal (corresponding to the dotted line in Fig. 1c) using two overlapping

frequency bands. The dashed line is a $1/\cos\theta$ curve characteristic of Bragg scattering from a Brillouin zone face. **b, $T(f, \theta)$** for a rotation about a five-fold rotation axis corresponding to the dashed line in Fig. 1c. Inset, schematic of the microwave horn and lens arrangement used for these measurements.

and D an integer) are required to describe the quasicrystal lattice. The icosahedral quasicrystal lattice of points can be constructed by projecting the points of a six-dimensional hypercubic lattice, the six-dimensional analogue of a three-dimensional cubic lattice. The coordinate of any lattice point can be described by the relation: $\mathbf{R} = \sum_{i=1}^6 n_i \mathbf{a}_i$, where the n_i are a subset of the integers and $\mathbf{a}_i$ are the six basis vectors: $\mathbf{a}_1 = (1, \tau, 0)$, $\mathbf{a}_2 = (-1, \tau, 0)$, $\mathbf{a}_3 = (0, 1, \tau)$, $\mathbf{a}_4 = (0, -1, \tau)$, $\mathbf{a}_5 = (\tau, 0, 1)$, $\mathbf{a}_6 = (\tau, 0, -1)$, and $\tau = (\sqrt{5} - 1)/2$, the golden mean. The structure has twelve five-fold, fifteen three-fold and thirty two-fold symmetry axes. We generate the lattice points of the icosahedral structure and create a solid structure by using equal length rods to connect points in pairs. We have made the overall shape a dodecahedron, so that each of the 12 outside faces will be perpendicular to a five-fold rotation axis, as shown in Fig. 1a.

Our crystals were created with a stereolithography machine (model SLA-250 from 3D Systems) that produces a solid plastic model by ultraviolet laser photopolymerization. The resolution is 0.1 mm lateral and 0.15 mm vertical. Both of our crystals have vertices connected with rods, of length $d = \sqrt{3}a/4 = 1$ cm. The rod diameter is 0.15 cm for our quasicrystal and 0.4 cm for the diamond structure. Our quasicrystal has 694 cells, 4,000 rods, and consists of 17.3 vol.% polymer. Our diamond structure has 500 cells and is 7.36% polymer. We measured the refractive index $\tilde{n}$ of the polymerized SLA5170 resin by placing a solid block in a waveguide and recording the transmission and reflection spectrum. For 33-GHz microwaves ($\lambda = 0.91$ cm in air), $\tilde{n} = 1.65 - 0.025i$. The resulting $(1/e)$ absorption length is 12 wavelengths. (The finite absorption from the polymer reduces the transmission approximately as $\exp(-2\omega\eta L/c)$, where $\eta = 0.025$ is the imaginary part of the refractive index and L is the transmission path length. The actual attenuation will depend on the geometry and the modes. In all curves in Figs 2–4 we have multiplied by the same simple exponential factor to reduce the background slope. This has no effect on the gap determination.)

Transmission measurements were made with a HP Model 8510C Vector Analyser in three bands, from 8 to 15, from 15 to 26 and from 26 to 42 GHz. To approximate plane waves, a single TE_{10} mode was coupled through two sets of horn-attached waveguides with two custom-made polystyrene microwave lenses as schematically shown in the inset in Fig. 2. Before the sample was inserted, the transmission spectrum of the set-up is recorded for normalization. The sample has different symmetry and dimensions in different directions, so the transmission spectrum should be sensitive to orientation and polarization. The sample was aligned so that the incident beam

was perpendicular to one of the sample's rotational symmetry axes. We rotated the sample along that rotational symmetry axis, and recorded the relative transmission spectrum every 2 degrees. For the quasicrystal, a rotation about the two-fold axis covers all the external

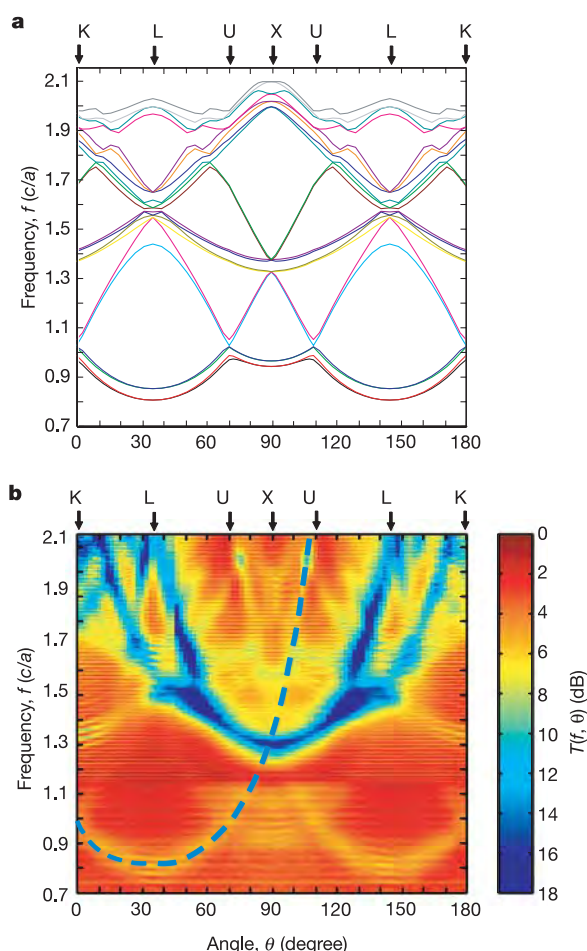


Figure 3 | Comparison of calculated bands and measured transmission for a diamond structure. a, Calculated dispersion relation f on the boundary of the first Brillouin zone versus θ , for the diamond structure along the dotted curve in Fig. 1d. b, $T(f, \theta)$ for the sample rotation along the same curve. There is excellent agreement at the photonic gap centre frequencies.

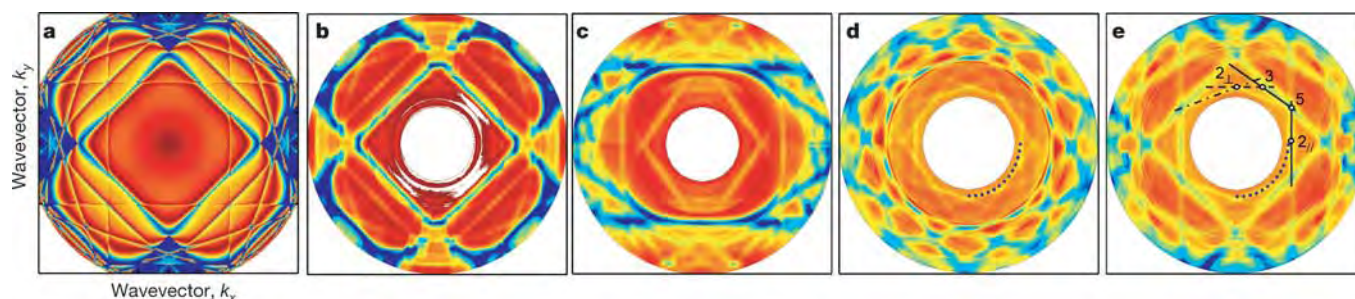


Figure 4 | Imaging of Brillouin zone for diamond and icosahedral quasicrystal structures. **a**, Brillouin zone for the diamond structure along the four-fold direction as seen in the contour plot of calculated frequency deviation ($\delta f = f - (c/\bar{n})|k|/(2\pi)$) versus k . **b–e**, The Brillouin zone can be seen in a plot of the measured $T(r=f, \theta=\theta)$ (using the same scale as Fig. 3) for the diamond lattice along the four-fold (dashed in Fig. 1d) axis (**b**) and

two-fold (dotted in Fig. 1d) axis (**c**); and for the quasicrystal along the five-fold (dashed in Fig. 1c) and two-fold (dotted in Fig. 1c) axes (**d**). The inner decagon in **d** and the solid and dashed lines in **e** correspond to the dashed and dotted lines in Fig. 1c. The dash-dotted line is a non-triacontahedral zone face.

points of the irreducible Brillouin zone and both polarizations. The region covered by the rotation about the two-fold (or five-fold) axis is shown as the dotted (dashed) red line in Fig. 1c. In Fig. 2 we show the measured transmission $T(f, \theta)$ for this rotation in overlapping plots from two frequency bands.

The simplest way to check our entire procedure is to perform the experiments on our diamond structure, where we are able to make a direct comparison with calculated stop gaps. We obtain the photonic bands of our rod-decorated diamond lattice, using the MIT Photonic Bands package¹³. Figure 3a shows the calculated band structure along a rotation about a two-fold axis. The rotation path is illustrated by the dotted red line in Fig. 1d. We found excellent agreement between the observed and calculated gap positions.

To gain insight into these complex spectra, we consider that gaps result from Bragg scattering. A wavevector that resides on the plane defined by a reciprocal lattice vector $\mathbf{G}$ is Bragg-scattered by $\mathbf{G}$. Such a wavevector satisfies the condition $\mathbf{k} \cdot \mathbf{G} = |\mathbf{G}|^2/2$ or equivalently, $|\mathbf{k}| = |\mathbf{G}|/(2\cos\theta)$. To lowest order, the centre frequency of a stop gap is therefore $f_G = (c/\bar{n})|\mathbf{G}|/(4\pi\cos\theta)$, where c is the speed of light in vacuum and $\bar{n}$ is the Bruggeman effective medium index¹⁴. The dashed curves in Figs 2a and 3b correspond to a $1/\cos(\theta - \theta_0)$ angular dependence consistent with Bragg scattering.

Compared with the diamond structure, the quasicrystal spectrum appears surprisingly less complex. Because the scattering function for a quasicrystal is a dense set of Bragg spots (of zero measure), we might have expected many gaps and zone faces to be intersecting. Instead there appear to be a few well-defined $1/\cos(\theta - \theta_0)$ curves in Fig. 2 and therefore few zone boundaries with sizeable gap formation.

Our method for visualizing the effective Brillouin zone structure is to invert the process by using the gaps to find the zone faces. We locate the points in reciprocal space responsible for the gaps by assuming $|\mathbf{k}(\theta)| \approx \bar{n}f(\theta)/(2\pi c)$. Then, in Fig. 4 we make polar plots of $T(r=f, \theta=\theta)$. For the diamond lattice, data from a rotation about a four-fold axis (dashed line in Fig. 1d) and a two-fold axis (dotted line in Fig. 1d) are shown in Fig. 4b and c, respectively. Figure 4a shows the calculated frequency deviation ($|\delta f| = |(f - (c/\bar{n})|k|)/(2\pi)|$) versus wavevector of the four-fold rotation.

Transmission data for our quasicrystal is shown in Figs 4d and e. The fact that the low-transmission regions correspond to straight lines indicates that the gaps lie on planes. These transmission polar plots, without any further analysis, directly give us the scattering planes and the effective Brillouin zones. In the smallest zone in Fig. 4d and 4e, we see the decagon from the five-fold rotation, and the additional symmetry planes from the orthogonal two-fold rotation, which correspond to the respective cuts of the triacontahedral Brillouin zone shown in Fig. 1c. (The wavevector corresponding to the edge centre of the smallest visible decagon in Fig. 4d is $\tau^2/\sqrt{\tau^2+1}$

in units of $2\pi/d$, where d is the rod length.) There are however, several unexpected features: a strong scattering plane along a 45° direction, the absence of strong scattering from the ' $2_\perp$ ' plane (dashed line in Fig. 4e; this may be a polarization effect due to the rod decoration of the unit cells), and another strong scattering plane (dash-dotted line in Fig. 4e) not on the triacontahedron. Note that a complete photonic bandgap would result if the dotted blue curves (cuts of a constant-frequency sphere) were contained within the gap of the zone boundary.

We find the following results. First, there is a relatively well-defined effective Brillouin zone, all of whose faces are consistent with the quasicrystal Bragg pattern. Second, the Brillouin zone structure is surprisingly simple despite the fact that a quasicrystal has a dense set of Bragg spots. Third, which is a key result for photonics, the measured Brillouin zone is close to spherical, with the largest difference in gap centre corresponding to 17% (dotted curve in Fig. 2a). Also, our experiments demonstrate that three-dimensional quasicrystals exhibit sizeable stop gaps on reasonably well-defined effective Brillouin zone faces. Hence, despite the quasiperiodicity, much of the intuition built up for conventional crystals may be applied. This experience with crystals suggests that our quasicrystal is far from optimized because it consists solely of thin rods connecting lattice points. A smoother, more spherical, multiply connected, unit-cell decoration with a more equal filled/void ratio would reduce polarization effects and enhance the gap overlap while maintaining the nearly spherical Brillouin zone. Laser tweezers used for particle trapping or two-photon polymerization would allow the construction of a quasicrystalline matrix of dielectric components with a photonic bandgap in the visible spectrum.

Received 18 May; accepted 27 June 2005.

- Joannopoulos, J., Meade, R. D. & Winn, J. *Photonic Crystals* (Princeton Univ. Press, Princeton, New Jersey, 1995).
- Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
- Forest, J. S. et al. Photonic-bandgap microcavities in optical waveguides. *Nature* **390**, 143–145 (1997).
- Chan, Y. S., Chan, C. T. & Liu, Z. Y. Photonic band gaps in two dimensional photonic quasicrystals. *Phys. Rev. Lett.* **80**, 956–959 (1998).
- Dal Negro, L. et al. Light transport through the band-edge states of Fibonacci quasicrystals. *Phys. Rev. Lett.* **90**, 055501 (2003).
- Cheng, S. S. M., Li, L., Chan, C. T. & Zhang, Z. Q. Defect and transmission properties of two dimensional quasiperiodic photonic band-gap systems. *Phys. Rev. B* **59**, 4091–4099 (1999).
- Kaliteevski, M. A. et al. Diffraction and transmission of light in low-refractive index Penrose-tiled photonic quasicrystals. *J. Phys: Condens. Matter* **13**, 10459 (2001).
- Schechtman, D., Blech, I., Gratias, D. & Cahn, J. W. Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.* **53**, 1951–1953 (1984).

9. Levine, D. & Steinhardt, P. J. Quasicrystals: a new class of ordered structures. *Phys. Rev. Lett.* **53**, 2477–2480 (1984).
10. Steinhardt, P. J. & Ostlund, S. *The Physics of Quasicrystals* (World Scientific, Singapore, 1987).
11. Pierce, F. S., Poon, S. J. & Biggs, B. D. Band-structure gap and electron transport in metallic quasicrystals and crystals. *Phys. Rev. Lett.* **70**, 3919–3922 (1993).
12. Fujiwara, T. & Tsuentsugu, H. in *Quasicrystals: The State of the Art* (eds DiVincenzo, D. P. & Steinhardt, P. J.), 2nd edn 361–390 (World Scientific, Singapore, 1999).
13. Johnson, S. G. & Joannopoulos, J. D. Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt. Express* **8**, 173–190 (2001).
14. Zeng, X. C., Bergman, D. J., Hui, P. M. & Stroud, D. Effective-medium theory for weakly nonlinear composites. *Phys. Rev. B* **38**, 10970–10973 (1988).

Acknowledgements We thank N. Jarosik for his extensive help on microwave measurements. We thank I. Aksay and the Chemical Engineering Department for the use of their SLA system, and the Gravity Group of the Princeton Physics Department for use of their microwave transmission measurement equipment. We also thank O. Crisafulli and R. Yang for help in the numerical aspects. This research was supported by NASA, by the US Department of Energy and by the National Science Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to P.J.S. (steinh@princeton.edu).

Liquid crystal 'blue phases' with a wide temperature range

Harry J. Coles¹ & Mikhail N. Pivnenko¹

Liquid crystal 'blue phases' are highly fluid self-assembled three-dimensional cubic defect structures that exist over narrow temperature ranges in highly chiral liquid crystals¹. The characteristic period of these defects is of the order of the wavelength of visible light, and they give rise to vivid specular reflections² that are controllable with external fields^{3–8}. Blue phases may be considered as examples of tuneable photonic crystals⁹ with many potential applications. The disadvantage of these materials, as predicted theoretically and proved experimentally¹, is that they have limited thermal stability: they exist over a small temperature range (0.5–2 °C) between isotropic and chiral nematic (N*) thermotropic phases, which limits their practical applicability. Here we report a generic family of liquid crystals that demonstrate an unusually broad body-centred cubic phase (BP I*) from 60 °C down to 16 °C. We prove this with optical texture analysis, selective reflection spectroscopy, Kössel diagrams and differential scanning calorimetry, and show, using a simple polarizer-free electro-optic cell, that the reflected colour is switched reversibly in applied electric fields over a wide colour range in typically 10 ms. We propose that the unusual behaviour of these blue phase materials is due to their dimeric molecular structure and their very high flexoelectric coefficients. This in turn sets out new theoretical challenges and potentially opens up new photonic applications.

There are three well-known¹ thermodynamically stable blue phases, BP III*, BP II* and BP I*, observed on cooling from the isotropic phase to the chiral nematic phase. BP III* is amorphous with a local cubic lattice structure in the director field, whereas BP II* and BP I* have a fluid three-dimensional periodic structure in the director field with simple cubic and body-centred cubic symmetry, respectively. For BP I* and BP II* the lattice periods are of the order of the wavelength of visible light and give rise to selective 'Bragg reflections'. These lead to potentially interesting photonic applications, such as three-dimensional blue-phase lasers¹⁰. Further, because of the fluidity of blue phases, external fields may be used to induce changes in the lattice parameters, thereby changing the specular reflection; this has led to simple colour change devices and optical filters. Hitherto, as predicted theoretically and observed experimentally, neat blue phases have only existed over a narrow temperature range a few degrees Celsius wide and this has limited such practical applicability. Several attempts have been made to widen the temperature interval of the blue phases, notably by polymer stabilization^{11,12}; the most recent report describes the stabilization of the three-dimensional cubic lattice in a defect confined polymer matrix¹³. Although an electro-optic Kerr effect, that is, a field-induced birefringence, was observed in an external electric field, this arose from the regions between the defect or disclination lines¹⁴, and the polymer lattice clearly restricted the deformation of the blue-phase lattice. Hence, no colour switching was observed.

In this paper, we describe novel blue-phase materials (BP I* and

BP II*) that are stable over a 40–50 °C temperature range, in which their reflectance band is switched linearly in an external field through deformation of the defect lattice, to give any desired reflectance colour at ambient temperatures. An electro-optic response with switching times of 10–40 ms and relaxation times of ~1–10 ms, depending on temperature, is also detected. We have so far made some 30 different mixtures that show blue phases 40–50 °C wide using both symmetric and non-symmetric bimesogens. The generic structure of our bimesogens is shown in Fig. 1a and for a typical blue-phase mixture of the type we describe here we use mixtures of the ratio 30.4% ($n = 7$), 35.1% ($n = 9$), 30.6% ($n = 11$) with 3.9% of the high twisted power (HTP) agent BDH1281 (available from Merck Chemicals and described in ref. 15). All concentrations are w/w and n refers to the number of methylene spacers in the alkyl chain linking the two mesogenic structures. The mixtures were studied by polarizing optical microscopy, light diffraction (Kössel diagrams), differential scanning calorimetry and electro-optic spectroscopy. The materials were contained in parallel plate glass cells with 7.0, 15 and 50- μm cell gaps and 4 mm $\times$ 4 mm indium-tin-oxide pixel electrodes. There were no alignment layers on the electrodes. The mixture showed the following phase sequence: isotropic 57.72 °C BP III*, 57.58 °C BP II*, 57.22 °C BP I*, 16.5 °C SmX* and -28 °C glass phase, where SmX* is an, as yet, unidentified smectic phase. The cooling rate was 0.001 °C min⁻¹ from the isotropic phase down to 56.84 °C to allow us to identify the phase transitions accurately, and then at 0.5 °C min⁻¹. Figure 1b, c, d shows the classical BP I* texture observed in our mixture on forming at 57.24 °C and then on cooling through 41 °C to 25 °C, some 32 °C below the BP II*–BP I* transition. To ensure that this was not a supercooled BP I*, we maintained the cell at 22–25 °C for over four months, and this texture remained unchanged irrespective of the sample thickness. Clearly the BP I* is identical at 41 °C and 25 °C and apart from some slight reflection-colour changes between 57.24 °C and 56.7 °C (see Fig. 2a), the platelet texture is identical. Further, on placing the sample between crossed polarizers and rotating the sample in plane the textures remain identical. This confirms that the reflected colours come from specular lattice reflections and not birefringence phenomena. We then indexed these lattice parameters using the Kössel diagram technique¹⁶. In Fig. 1e we give the Kössel diagrams and indexing for two specific platelets at 25 °C. These are identical to those recorded at higher temperatures up to the transition region of 56.7 °C. Differential scanning calorimetry measurements, on thermal cycling, confirmed the stability of these phases and the data are given in the Supplementary Fig. 1 along with textures for the BP I*, BP II* and BP III* phase transitions (Supplementary Fig. 2). Textures for the BP I* phase in a 50- μm -thick sample are given in Supplementary Fig. 3, confirming that the stability of the BP I* is not due to super cooling or due to sample thickness.

Given that the optical textures are classical blue phases over such a wide temperature range and that the lattice periodicities could be

¹Centre of Molecular Materials for Photonics and Electronics, Engineering Department, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK.

indexed using Kössel diagrams, we examined the spectral, thermo- and electro-optic properties in more detail. We repeated the same cycle of cooling from the isotropic phase through the BP II* and BP I* phases (Fig. 2a, b). BP II* gives a near-ultraviolet reflection which decreases in wavelength on cooling, commensurate with a lattice contraction. At the BP II*–BP I* transition the lattice parameter

increases to give a 60-nm increase in the reflection band, which then gradually increases over a 0.4 °C temperature range and then maximizes at ~500 nm (that is, green reflected light; Fig. 2a). On further cooling, over a ~40 °C range, this (200) reflection wavelength then very gradually decreased, by only 10 nm, until the SmX* phase is reached (Fig. 2b). We also monitored the (110) reflection from a 'red' platelet which showed exactly the same behaviour (Fig. 2b). The

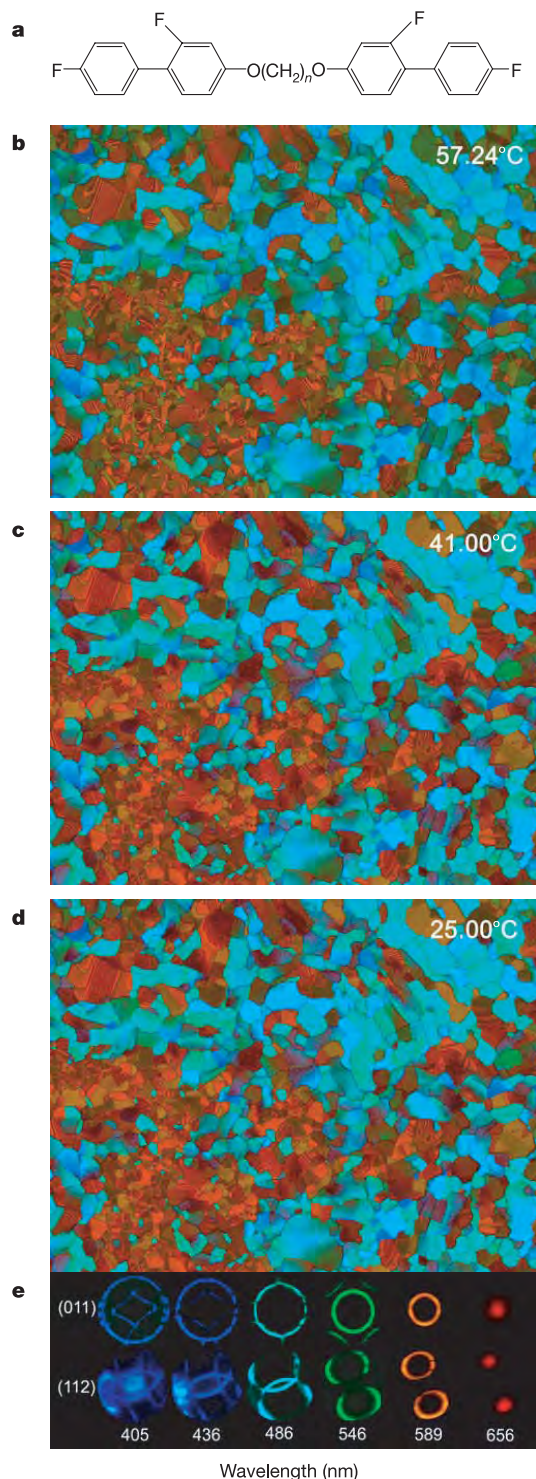


Figure 1 | Blue phase materials, temperature dependence of optical textures and Kössel diagrams. **a**, Generic chemical structure of the materials. **b–d**, Typical textures of the BP I* phase over a temperature range from 57 °C to 25 °C in a 7- μ m cell with cooling rate of 0.5 °C min⁻¹. **e**, The Kössel diagrams for the domains in (011) and (112) orientations for different wavelengths.

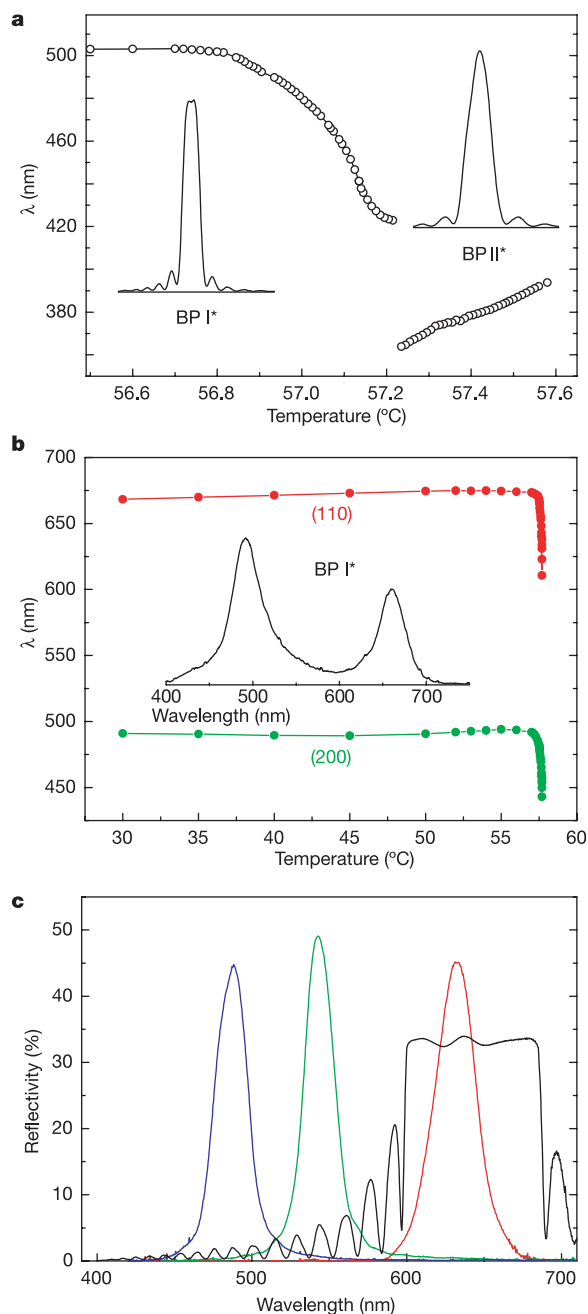


Figure 2 | Spectral properties of blue phases as a function of temperature and composition. **a**, Temperature dependence of the wavelength of the Bragg reflection for the BP I* and BP II* phases over a narrow temperature range and a typical shape of the reflection spectra (insert). **b**, Temperature dependence of the wavelength of the Bragg reflections for two domains of (110) and (200) orientation of the BP I* phase over a broad temperature range. **c**, Selective reflection spectra in the BP I* phase at 25 °C for mixtures with 3.9, 4.8 and 5.3% concentration of the chiral additive (red, green and blue spectral lines, respectively), and the selective reflection spectrum in the N* phase (black line) for a commercially available material E49 with the same chiral additive.

insets for Fig. 2a, b show the classical narrow linewidth associated with reflections from blue phases. By adjusting the concentration of chiral additive we altered the pitch of the system to give red, green and blue reflections in BP I; these are compared with a typical chiral nematic using the same chiral additive in Fig. 2c. All sets of data are at 25 °C and it is clear that the narrow spectral lines of the blue phases give much narrower linewidths and higher reflection intensities than the chiral nematic at the same temperature. From these data sets—that is, the optical textures, the Kössel diagrams, the differential scanning calorimetry and the spectral linewidths—it is clear that we observed blue phases over a 40–50-K-wide temperature range, in pure mixtures unaided by supercooling effects achieved using polymer networks¹³.

We then studied the electric-field dependency of the selective reflection in BP I* at 25 °C by applying increasing and then decreasing pulsed alternating current (a.c.) electric fields (100 Hz; Fig. 3a). This shows two very distinct regimes. For fields below $14 \text{ V } \mu\text{m}^{-1}$, with increasing field strength, there is a gradual decrease in reflected wavelength from $\sim 572.5 \text{ nm}$ down to 568 nm with a small residual hysteresis ($< 1 \text{ nm}$) between increasing and decreasing field. This is consistent with small changes in refractive index due to local director

orientation within the (110) lattice. However, above the critical or threshold field for these materials, the reflected colour changes rapidly from red/orange to green/blue (that is, 572 nm to 506 nm) as the field is increased to $18 \text{ V } \mu\text{m}^{-1}$. On removal of the field the green/blue reflected colour reverted back to red/orange. The hysteresis between increasing and decreasing field is very small (a few nanometres). This dramatic field-switchable colour is shown in Fig. 3b. As shown in Fig. 3a, the reflected wavelength is linearly dependent on field in this field regime (that is, between $14 \text{ V } \mu\text{m}^{-1}$ and $18 \text{ V } \mu\text{m}^{-1}$). These changes are quite clearly due to electric-field-induced lattice distortions: that is, electrostriction. Further, up to the start of the transition from BP I*–BP II* these wavelength changes are independent of temperature to within a few nanometres. The reflection images recorded in Fig. 3b are for typical pixel-sized indium-tin-oxide electrodes (that is, $50 \text{ } \mu\text{m} \times 50 \text{ } \mu\text{m}$) of a dot matrix array on glass substrates. On application of a pulsed a.c. field (100 Hz) the lattice distortion has a rise time of 53 ms (at 30 °C) in response to a $14 \text{ V } \mu\text{m}^{-1}$ (root mean square, r.m.s.) field and on removal of the field the decay or fall time back to the undisturbed lattice was 7 ms. As the temperature is increased for the same field these response times decrease on a logarithmic scale so that at $\sim 40 \text{ } ^\circ\text{C}$ the rise time is 20 ms and the fall time drops to $\sim 3 \text{ ms}$ (Fig. 3c).

To understand the origins of these blue-phase materials and their properties we consider the macroscopic behaviour of the bimesogens in more detail. We recently discovered¹⁷ that these bimesogens give rise to the highest recorded (an order of magnitude greater than for monomesogens) flexoelectro-optic ratios (e/k) in the chiral nematic phase (N^*). In the flexoelectro-optic effect¹⁸ a chiral nematic is constrained to lie in the plane of a simple electro-optic device, formed by the two glass substrates with indium-tin-oxide electrodes on each face, hybrid alignment layers and the N^* material positioned between the substrates. On application of an electric field, the optic axis rotates in the plane of the device by an angle φ defined by $\tan \varphi = \frac{eE}{kK}$, where e is the average of the flexoelectric coefficient, k is the average of the splay and bend elastic constant, K is the helical wave vector $2\pi/p$ (where p is helix pitch) and E is the applied field. The figure of merit here is e/k . In bimesogenic mixtures similar to those described herein¹⁹ we have now achieved switching angles of $\pm 86^\circ$, which implies a very high e/k for these bimesogenic materials. We have also measured high k values compared to ‘normal’ nematic liquid crystals, so these bimesogens can readily deform in the director field to give extraordinarily high values of e . It is the demonstration of the high e and k values associated with the bimesogens, when incorporated into short-pitch chiral structures, that led us to examine the blue phases formed by these materials. The large flexoelectric effect arises from the distortion of the rapidly varying director field in the chiral nematic phase and similar distortions must be present in BP I* and BP II* at the site of the line singularity ($s = -1/2$) formed by the intersection of the three orthogonal double twist cylinders characteristic of the blue phases¹. We believe that it is the large localized flexoelectric polarization generated by such director distortions close to these disclination line singularities that stabilizes the blue phases with the bimesogenic materials and effectively ‘pins’ the lattice defects. Macroscopically, owing to symmetry, the net polarization will be zero when averaged around the singularity.

Here we report naturally occurring broad-temperature-range blue-phase materials that demonstrate large analogue reflected-colour switching induced in electric fields with response times of a few milliseconds. The zero-field reflected colour is chosen by the concentration of HTP chiral additive (see Fig. 2c). The pixellated test device did not incorporate polarizers, analysers or colour filters and we believe that these materials will lead to a new generation of transreflective bright, low-power-consumption liquid crystal displays²⁰. Further, the use of voltage-controlled colour means that each pixel can reflect red, green or blue (RGB) using temporal dither and reduce the pixel and thin film transistor (TFT) density by a factor of three. The materials may also be used in tuneable optical filters.

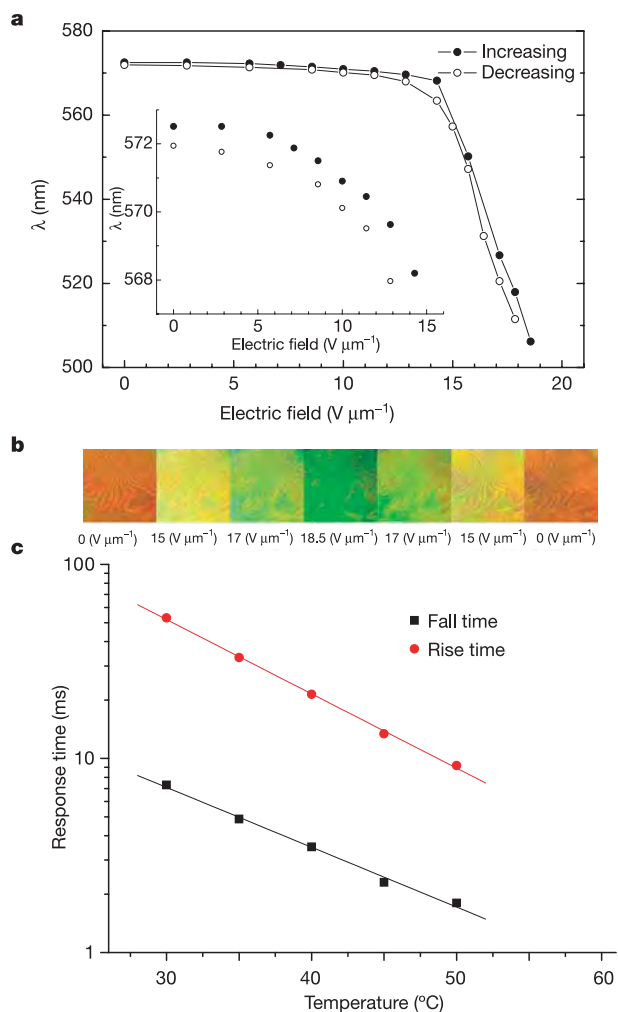


Figure 3 | Voltage- and time- dependent colour switching of BP I*. **a**, Electric-field dependence of the selective reflection peak for BP I* at 25 °C. **b**, Colour switching of an individual display pixel observed on reflection as a function of applied field in BP I at 25 °C. **c**, The temperature dependence of the rise and fall response times observed between crossed polarizers. The inset shows an expanded plot of the selective reflection as a function of the applied electric field at low fields (that is, below $14 \text{ V } \mu\text{m}^{-1}$). The axis labels are the same for both plots.

Further, because of the photonic bandgap nature of BPI, the materials may also readily be incorporated into three-dimensional organic lasers¹⁰ but with a wide temperature range of stability. The electric-field-induced lattice distortions will add a new generation of continuously tuneable laser sources and indeed opens up new perspectives for liquid-crystal-based photonics.

Received 3 March; accepted 8 June 2005.

1. Crooker, P. P. in *Chirality in Liquid Crystals* (eds Kitzerow, H.-S. & Bahr, C.) 186–222 (Springer, New York, 2001).
2. Stegemeyer, H. *et al.* Thermodynamic, structural and morphological studies on liquid-crystalline blue phases. *Liq. Cryst.* **1**, 3–28 (1986).
3. Gleeson, H. F., Simon, R. & Coles, H. J. Electric field effects and two frequency colour switching in the cholesteric and blue phases of nematic/cholesteric mixtures (Tenth Int. Liquid Crystal Conf. (York, UK, July 1984; Paper G17). *Mol. Cryst. Liq. Cryst.* **129**, 37–52 (1985).
4. Gerber, P. R. Electro-optical effects of a small pitch blue phase system. *Mol. Cryst. Liq. Cryst.* **116**, 197–206 (1985).
5. Coles, H. J. & Gleeson, H. F. Electric field induced phase transitions and colour switching in the blue phases of chiral nematic liquid crystals. *Mol. Cryst. Liq. Cryst.* **167**, 213–225 (1989).
6. Dmitrienko, V. E. Electro-optic effects in blue phases. *Liq. Cryst.* **5**, 847–851 (1989).
7. Kitzerow, H.-S. The effect of electric fields on blue phases. *Mol. Cryst. Liq. Cryst.* **202**, 51–83 (1991).
8. Heppke, G., Jerome, B., Kitzerow, H.-S. & Pieranski, P. Electrostriction of the cholesteric blue phases BPI BP II in mixtures with positive dielectric anisotropy. *J. Phys.* **50**, 2291–2298 (1991).
9. Etchegoin, P. Blue phases of cholesteric liquid crystals as thermotropic photonic crystals. *Phys. Rev. E* **62**, 1435–1437 (2000).
10. Cao, W., Muñoz, A., Palffy-Muhoray, P. & Taheri, B. Lasing in a three-dimensional photonic crystal of the liquid crystal blue phase II. *Nature Mater.* **1**, 111–113 (2002).
11. Kitzerow, H.-S. *et al.* Observation of blue phases in chiral networks. *Liq. Cryst.* **14**, 911–916 (1993).
12. Bohley, C. & Scharf, T. Blue phases as photonic crystals. *Proc. SPIE* **5184**, 202–208 (2003).
13. Kikuchi, H., Yokota, M., Hisakado, Y., Yang, H. & Kajiyama, T. Polymer-stabilized liquid crystal blue phases. *Nature Mater.* **1**, 64–69 (2002).
14. Hisakado, Y., Kikuchi, H., Nagamura, T. & Kajiyama, T. Large electro-optic Kerr effect in polymer stabilised liquid-crystalline blue phases. *Adv. Mater.* **17**, 96–98 (2005).
15. Parry, O., Nolan, P., Farrand, L. & May, A. L. Chiral dopants. UK Patent GB2329636 (31 March 1999).
16. Pieranski, P. in *Chirality in Liquid Crystals* (eds Kitzerow, H.-S. & Bahr, C.) 28–66 (Springer, New York, 2001).
17. Coles, H. J., Coles, M. J., Perkins, S. P., Musgrave, B. M. & Coates, D. Bimesogenic compounds and their use in flexoelectric liquid crystal devices. UK Patent GB2356629. (30 May 2001).
18. Patel, J. S. & Meyer, R. B. Flexoelectric electro-optics of a cholesteric liquid crystal. *Phys. Rev. Lett.* **58**, 1538–1540 (1987).
19. Coles, H. J., Clarke, M. J., Morris, S. M., Broughton, B. J. & Blatch, A. E. Strong flexoelectric behaviour in bimesogenic liquid crystals. *J. Appl. Phys.* (submitted).
20. Wu, S. T. & Yang, D.-K. *Reflective Liquid Crystal Displays* Ch. 1 (Wiley, Chichester, UK, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements H.J.C. thanks the EPSRC, UK, who funded the materials research under their Displays and Functional Materials initiatives.

Author Contributions H.J.C. invented the flexoelectric materials, was responsible for the project planning, wrote the Letter and was responsible for the tentative explanation. M.N.P. carried out all the experimental work, including microscopic texture and data analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.J.C. (hjc37@cam.ac.uk).

Short-term variations in the oxidizing power of the atmosphere

Martin R. Manning¹, David C. Lowe², Rowena C. Moss², Gregory E. Bodeker² & William Allan²

The hydroxyl radical is the predominant atmospheric oxidant¹, responsible for removing a wide range of trace gases, including greenhouse gases, from the atmosphere. Determination of trends and variability in hydroxyl radical concentrations^{2,3} is critical to understanding whether the 'cleansing' properties of the atmosphere are changing. The variability in hydroxyl radical concentrations on annual to monthly timescales, however, is difficult to quantify. Here we show records of carbon monoxide containing radiocarbon (¹⁴CO), which is oxidized by hydroxyl radicals^{4,5}, from clean-air sites at Baring Head, New Zealand, and Scott Base, Antarctica, spanning 13 years. Using a model study, we correct for known variations in production of ¹⁴CO (refs 6, 7), allowing us to exploit this species as a diagnostic for short term changes in hydroxyl radical concentrations. We find no significant long-term trend in hydroxyl radical concentrations but provide evidence for recurring short-term variations of around ten per cent persisting for a few months. We also find decreases in hydroxyl radical concentrations of up to 20 per cent, apparently triggered by the eruption of Mt Pinatubo in 1991 and by the occurrence of extensive fires in Indonesia in 1997.

¹⁴CO is produced in the atmosphere by cosmic-ray-induced neutrons and has an atmospheric lifetime of about three months. Extensive measurements of ¹⁴CO became practical after technical

advances^{8–11} in the late 1980s. Here we consider the two longest records of ¹⁴CO based on these methods, which are from Southern Hemisphere clean-air sites at Baring Head (41.4°S), New Zealand, and Scott Base (77.8°S), Antarctica, and cover 1989 to 2003. Our focus is on the ¹⁴CO produced directly from ¹⁴C, so we remove a 'recycled' fraction¹² due to surface emissions of CO containing ¹⁴C and oxidation of atmospheric ¹⁴CH₄ to ¹⁴CO (see Methods). These corrections typically amount to 20% of the measured values and result in a 'primary' ¹⁴CO concentration directly attributable to recent ¹⁴C production.

Figure 1 shows primary ¹⁴CO concentrations from our two sites and three features are evident. First, there is a large seasonal cycle of about ±40% around the annual mean owing to the strong seasonality of hydroxyl radicals (OH) at these latitudes together with the sensitivity of ¹⁴CO to OH variations. For comparison, the seasonal cycle of the longer-lived methylchloroform species, also used to diagnose OH, varies by about ±3% around its annual mean. Second, annual mean ¹⁴CO concentrations varied by more than 50%, reflecting solar modulation of ¹⁴C production⁶ during the last Schwabe cycle and apparently following independently estimated ¹⁴C production rates⁷ for the period. Third, despite their large latitudinal separation, concentrations at the two sites generally agree to within 1 molecule cm⁻³, although in some years Antarctic

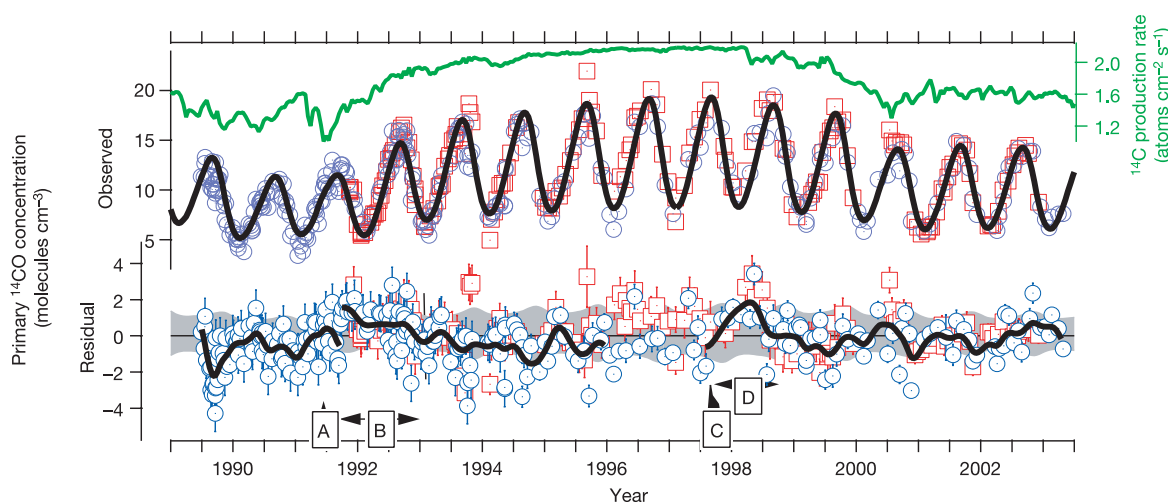


Figure 1 | ¹⁴CO data and simulation. The upper section shows primary ¹⁴CO concentrations for individual samples from Baring Head (blue circles) and Scott Base (red squares) from 1989 to 2003. The green and black curves show ¹⁴C production rates (right-hand scale) and the best-fit simulation of ¹⁴CO discussed in the text. The lower section shows residuals (observed minus simulated), measurement uncertainties (1-σ), and a Reinsch spline smooth curve, fitted separately to residuals before and after the discontinuity in

September 1991, and masked out during 1996 and 1997—when the sites differ as discussed in the text. The grey shaded region shows the deviations from the best fit that would be caused by a 10% reduction or increase in OH concentrations. A, time of eruption of Mt Pinatubo; B, period of high Southern Hemisphere CH₄ concentrations²²; C, time profile of Indonesian fires²⁴; and D, period of high Southern Hemisphere CO concentrations²⁶.

¹IPCC Working Group I Support Unit, Boulder, Colorado 80305, USA. ²National Institute of Water and Atmospheric Research, Wellington 6003, New Zealand.

values are slightly higher around the time of the October seasonal maximum. Similar very low ^{14}CO latitudinal gradients south of 40°S have been reported previously from ship sampling¹³ and aircraft sampling in the troposphere¹⁰.

The latitudinal structure of primary ^{14}CO is further elaborated in Fig. 2, showing latitudinal profiles from recent ship sampling at different times of the year and normalized to a constant ^{14}C production rate. Two distinct regimes can be identified: a tropical region with low ^{14}CO at all times of the year; and an extra-tropical region with very small latitudinal gradient and a strong seasonal cycle.

Chemical tracer models can reproduce most of the observed patterns of ^{14}CO variation^{14,15} but do not fully explain the absence of latitudinal gradients in the extra-tropical southern hemisphere (ETSH). For example, a recent study¹⁵ shows winter-time differences in ^{14}CO of about 4 molecules cm^{-3} between 40°S and 70°S for typical ^{14}C production rates of 2 molecules $\text{cm}^{-2}\text{s}^{-1}$, significantly larger than observed differences. We believe that very low gradients in the ETSH are caused partly by rapid mixing over the Southern Ocean but also by a fortuitous effect of the spatial patterns of ^{14}CO production, cross-tropopause transport, and oxidation within the region.

Here we focus on explaining the temporal variations of ^{14}CO in the region between New Zealand and Antarctica by analysing the factors that control concentrations there. The lack of latitudinal gradient implies that the net effects of north-south transport are minor within the region, so the uniform concentration is controlled predominantly by the combination of *in situ* production in the troposphere, production in the stratosphere followed by stratosphere–troposphere exchange (STE), and removal by OH in the troposphere. A two-box model is used to quantify these factors (see Methods).

Our initial simulation uses *a priori* estimates of ^{14}C production⁷, OH concentrations¹⁴, and Southern Hemisphere STE¹⁶ from previous studies. Reaction with OH is assumed to be the only removal process in the troposphere, and the net stratospheric removal rate is assumed to be equal to that in the troposphere. This produces an excellent fit to our observations, explaining more than 99.9% of their variance.

We then consider optimization of our model by scaling and shifting the *a priori* production and removal terms, varying the seasonal phasing of the OH and STE cycles, and the ratio of

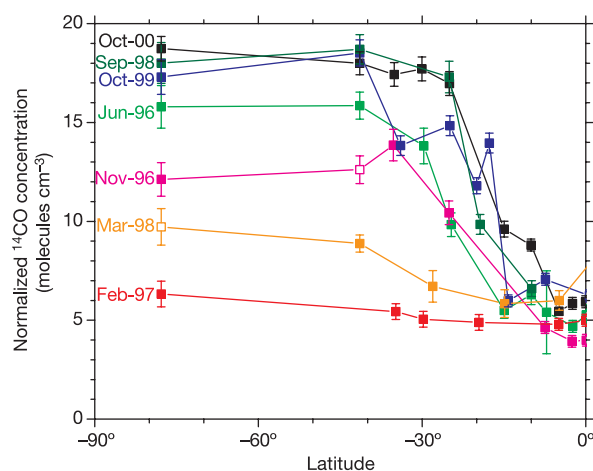


Figure 2 | Latitudinal profiles of primary ^{14}CO concentrations. Data are shown from seven ship voyages in the Pacific³⁰, together with contemporaneous data from Baring Head and Scott Base. Two hollow symbols denote points that have been interpolated from data in adjacent months at the fixed sites. Concentrations have been scaled to a constant ^{14}C production rate of 2 atoms $\text{cm}^{-2}\text{s}^{-1}$, and are shown with $1\text{-}\sigma$ measurement uncertainties.

stratospheric to tropospheric removal rates (see Methods). The most significant change caused by optimization is to move the annual OH maximum from 1 January to 17 December, close to the time of maximum solar irradiance. Figure 1 shows the optimal fit and the corresponding residuals.

Because the best fit to our data requires only small adjustments to the *a priori* estimates of ^{14}CO production and removal, we believe this provides independent confirmation of the specific chemistry and physics process models used to derive those estimates. For example, two alternative estimates of ^{14}C production with different systematic behaviour through the last solar cycle are inconsistent with our data. (See Methods and Supplementary Material for details of this and other constraints imposed by our data on model terms.)

The residuals shown in Fig. 1 exhibit temporally coherent behaviour including abrupt shifts, such as an apparent discontinuity in September 1991, and persistence of positive or negative values. Statistical tests confirm that these residuals are highly non-random, suggesting that they are caused, at least in part, by anomalies in the processes controlling ^{14}CO concentrations. Candidates for such anomalies include changes in intra-hemispheric exchange, anomalous ^{14}C production, or departures of either STE or OH from the repeating seasonal cycles used in our model.

Figure 2 implies that any variations in exchange between the tropics and extra-tropics would affect Baring Head data to a greater degree than Scott Base data. As seen in Fig. 1, differences did occur between these sites during 1996 and 1997 with lower values at Baring Head suggesting enhanced exchange with the tropics. However, for most of 1989–2003, close agreement between the sites suggests that intrahemispheric transport anomalies play a minor role.

The largest anomalies in ^{14}C production are expected to be from solar proton events, which can cause large increases in production for short periods. Model analysis¹⁷ of the effect of such events on tropospheric ^{14}CO suggests that the largest event during the period considered would increase Baring Head concentrations by less than 10%. Furthermore, there appears to be no relationship between the timing of the eight largest solar proton events in the period and observed ^{14}CO anomalies, so production anomalies do not explain our results.

There is observational evidence for both episodic behaviour and trends in atmospheric circulation in the upper-troposphere–lower-stratosphere region that could affect STE. This includes downward propagation of stratospheric dynamical features affecting the troposphere¹⁸ and increases in tropopause height¹⁹, which have been linked to climate change²⁰. To consider whether such dynamical effects might influence our data, we determined anomalies in potential vorticity averaged over 30°S to 70°S near the tropopause from re-analysed meteorological data²¹ and found that, with the exception of the Pinatubo eruption event, their timing does not match those of observed ^{14}CO anomalies.

Further evidence for lack of a strong STE influence on tropospheric ^{14}CO is provided by the absence of any anomaly in late 2002 during a period of extraordinary behaviour of the Antarctic polar vortex²². This observational evidence for weak sensitivity to STE anomalies is consistent with our model analysis, in which only about 15% of the ^{14}CO observed in the ETSH troposphere comes from the stratosphere (other models give a wide range for this fraction¹⁵—see Supplementary Material).

Thus, we hypothesize that anomalies in tropospheric OH are the principal cause of ^{14}CO anomalies that occur simultaneously at both Baring Head and Scott Base. To indicate the scale of variations that are implied, Fig. 1 shows the ^{14}CO deviations that would be caused by a 10% increase or decrease in OH. On this basis, we infer an abrupt OH decrease in September 1991 of 15% to 20% followed by gradual recovery over the next six to twelve months (period B in Fig. 1). A second large anomaly starts around September 1997 and extends to mid-1998, when there was a rapid recovery (period D in Fig. 1). Apart from these cases it appears that there is more or less continual

variability of OH in the order of $\pm 10\%$ around its climatological mean.

The first large ^{14}C anomaly occurs shortly after the Mt Pinatubo eruption, which injected large amounts of SO_2 into the lower stratosphere, reducing the flux of ultraviolet radiation into the troposphere, and decreasing OH production. A previous study²³ estimated a 7 to 8% reduction in OH due to reduced ultraviolet flux but did not quantify the effect of other amplifying factors such as an observed reduction in atmospheric water vapour. An OH reduction of order 10% decaying over about one year is consistent with anomalously high CH_4 growth rates in this period.

The second major anomaly follows extensive biomass burning in the Indonesian region²⁴. Regional CO concentrations in the upper troposphere reached 2.5 times their normal values²⁵ and CO concentrations remained anomalously high in the ETSH for about one year (ref. 26). This will have decreased OH, because CO is its principal removal agent, but there are no estimates of the magnitude of that effect. Thus the direction and timescales of the two major anomalies seen in ^{14}C are consistent with independent but less quantitative evidence for changes in atmospheric oxidation rates, supporting our hypothesis that OH is also the principal cause of other temporal anomalies in ^{14}C .

Over the 13-year period our data are consistent with analyses of methylchloroform data², implying no significant trend in Southern Hemisphere OH during the 1990s. Although higher-than-average OH values are indicated at the beginning of our data record, our analysis may be sensitive to end-effects or we may have observed the end of a transient increase, so we would not infer any long-term trend from our data. It should also be noted that inferring small trends in OH from ^{14}C data would require more careful consideration of possible trends in STE and tropopause height.

The OH variability indicated by our analysis has implications for the accuracy with which trace gas emissions can be determined from atmospheric concentrations using inversion techniques that ignore interannual variability in OH, and hence in removal rates. Although the type of OH variability shown here will have smaller effects on trace gases with longer lifetimes, we believe there is a need for caution when using inverse techniques to compare source emissions from year to year, even for species such as CH_4 .

METHODS

Observations. Our data are corrected for a recently quantified artefact caused by the formation of ^{14}C in sample storage cylinders between the times of collection and analysis²⁷. These corrections are up to 3 molecules cm^{-3} and significantly reduce an apparent gradient between New Zealand and Antarctica. A recent ^{14}C climatology¹² showed such a gradient, but recognized the need to consider storage correction terms that were not quantified at that time.

The recycled fraction of ^{14}C , due to surface CO sources and CH_4 oxidation, is estimated by multiplying a model-derived $^{14}\text{C}/^{12}\text{C}$ ratio for clean air CO by measured CO concentrations for each sample. A recent inversion study²⁸, which fitted CO sources to observations, is used to estimate the amount of CO due to each source type at each site. Typical $^{14}\text{C}/^{12}\text{C}$ ratios for each source type are then used to determine $^{14}\text{C}/^{12}\text{C}$ ratios for total CO. To prevent undue extrapolation, samples whose CO concentration exceeds the modelled clean air value by more than 50% are excluded from further analysis. The recycled ^{14}C fraction is in the range 10% to 20% and its uncertainties are small relative to measurement errors.

Analysis of the a priori case. Mass balance constraints on tropospheric ^{14}C in the ETSH are approximated by a two-box model incorporating production, loss and exchange terms in stratospheric and tropospheric boxes using the differential equations:

$$\begin{aligned} M_S \frac{dC_S}{dt} &= -L_S M_S C_S(t) + P_S + X_{ST}(C_T(t) - C_S(t)) \\ M_T \frac{dC_T}{dt} &= -L_T M_T C_T(t) + P_T + X_{ST}(C_S(t) - C_T(t)) \end{aligned} \quad (1)$$

where subscripts S and T denote stratosphere and troposphere, M is the air mass, C is the ^{14}C mass mixing ratio, L is the loss rate, P is the production rate, and X_{ST} is the rate of air mass exchange between stratosphere and troposphere.

Further approximations are:

$$P_S = 0.6P_A, P_T = 0.4P_A, L_S = fL_T, \text{ and } M_T = 9M_S$$

where P_A is the total ^{14}C production in the atmosphere and the 60% fraction assigned to the stratosphere is a mid-range value from different estimates²⁹. f is a fixed ratio between stratospheric and tropospheric loss rates, initially assumed to be one.

The ^{14}C production rate, P_A , for the 13-year period is based on ^{14}C production rates recommended by ref. 7 and derived from neutron count rates extending an earlier *ab initio* approach⁶. Of the various approaches to estimating ^{14}C production¹², this avoids assumptions about the time lag for solar influences to propagate through the heliosphere and uses a more direct measure of the agent producing ^{14}C . We assume a 90% yield of ^{14}C from ^{14}C oxidation⁴.

The seasonally varying loss rate, L_T , is the product of the pressure-dependent rate constant for the CO + OH reaction and OH concentrations from a global atmospheric chemistry model¹⁴ that is consistent with observations of several species linked to OH. The seasonally varying exchange term X_{ST} is taken from a study¹⁶ of the effect of STE anomalies on tropospheric trace gas concentrations and based on dynamical studies of atmospheric circulation. Further details are provided in the Supplementary Information.

Optimization of parameters. To explore the constraints imposed by our ^{14}C data on this model we rewrite:

$$\begin{aligned} L_T(t) &= l_0 L_T^{\text{Avg}} + l_1 (L_T^0(t - w_L) - L_T^{\text{Avg}}) \\ P_A(t) &= p_0 P_A^{\text{Avg}} + p_1 (P_A^0(t) - P_A^{\text{Avg}}) \\ X_{ST}(t) &= x_0 X_{ST}^{\text{Avg}} + x_1 (X_{ST}^0(t - w_X) - X_{ST}^{\text{Avg}}) \end{aligned} \quad (2)$$

where $L_T^0(t)$, $P_A^0(t)$ and $X_{ST}^0(t)$ are the a priori model terms defined above and L_T^{Avg} , P_A^{Avg} , X_{ST}^{Avg} are their average values. Dimensionless parameters l_0 , p_0 and x_0 adjust average values of the new model terms; l_1 , p_1 and x_1 adjust their range of variation; and w_L and w_X are time-lag parameters with units of time. Together with f defined in equation (1), this provides nine parameters that can be varied to provide a least-squares fit to the pooled data from both sites. The a priori model terms are obtained when w_L and w_X are zero and all other parameters are one.

The effect of changes in production on the mixing ratios in equation (1) can be compensated by corresponding changes in loss rate¹⁵ causing the full optimization outlined in equation (2) to be very poorly determined. Thus we fix $p_0 = 1$, effectively determining average loss rates relative to the average a priori production term. We also introduce bayesian parameter constraints by combining the sum of squared error weighted data residuals with the sum of squares of parameter changes relative to their a priori uncertainties, which are taken from the literature sources for the various model terms (see Supplementary Information).

The most significant parameter changes are to shift w_L to -15 days, reduce l_1 by 7%, and increase p_1 by 13%. However, the last two parameters are negatively correlated—indicating that the data are consistent with smaller joint changes in each.

Further details, including a determination of a posteriori parameter uncertainties, are provided in the Supplementary Information.

Received 26 April; accepted 7 June 2005.

- Levy, H. II Normal atmosphere: Large radical and formaldehyde concentrations predicted. *Science* **173**, 141–143 (1971).
- Prinn, R. G. *et al.* Evidence for substantial variations of atmospheric hydroxyl radicals in the past two decades. *Science* **292**, 1882–1888 (2001).
- Krol, M. & Lelieveld, J. Can the variability in tropospheric OH be deduced from measurements of 1,1,1-trichloroethane (methyl chloroform)? *J. Geophys. Res.* **108**, 4125, doi:10.1029/2002JD002423 (2003).
- Weinstock, B. & Niki, H. Carbon monoxide balance in nature. *Science* **176**, 290–292 (1972).
- Volz, A., Ehhalt, D. H. & Derwent, R. G. Seasonal and latitudinal variation of ^{14}C and the tropospheric concentration of OH radicals. *J. Geophys. Res.* **86**, 5163–5171 (1981).
- Masarik, J. & Beer, J. Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J. Geophys. Res.* **104**, 12099–12111 (1999).
- Lowe, D. C. & Allan, W. A simple procedure for evaluating global cosmogenic ^{14}C production in the atmosphere using neutron monitor data. *Radiocarbon* **44**, 149–157 (2002).
- Brenninkmeijer, C. A. M. *et al.* Interhemispheric asymmetry in OH abundance inferred from measurements of atmospheric ^{14}C . *Nature* **356**, 50–52 (1992).

9. Brenninkmeijer, C. A. M. Measurement of the abundance of ^{14}CO in the atmosphere and the $^{13}\text{C}/^{12}\text{C}$ and $^{18}\text{O}/^{16}\text{O}$ ratio of atmospheric CO with applications in New Zealand and Antarctica. *J. Geophys. Res.* **98**, 10595–10614 (1993).
10. Mak, J. E., Brenninkmeijer, C. A. M. & Tamareis, J. Atmospheric ^{14}CO observations and their use for estimating carbon monoxide removal rates. *J. Geophys. Res.* **99**, 22915–22922 (1994).
11. Brenninkmeijer, C. A. M., Lowe, D. C., Manning, M. R., Sparks, R. J. & van Velthoven, P. F. J. The ^{13}C , ^{14}C and ^{18}O isotopic composition of CO, CH_4 , and CO_2 in the higher southern latitudes lower stratosphere. *J. Geophys. Res.* **100**, 26163–26172 (1995).
12. Jöckel, P. & Brenninkmeijer, C. A. M. The seasonal cycle of cosmogenic ^{14}CO at the surface level: A solar cycle adjusted, zonal-average climatology based on observations. *J. Geophys. Res.* **107**, 4656, doi:10.1029/2001JD001104 (2002).
13. Quay, P. *et al.* Atmospheric ^{14}CO : a tracer of OH concentration and mixing rates. *J. Geophys. Res.* **105**, 15147–15166 (2000).
14. Spivakovsky, C. M. *et al.* Three-dimensional climatological distribution of tropospheric OH: Update and evaluation. *J. Geophys. Res.* **105**, 8931–8980 (2000).
15. Jöckel, P., Brenninkmeijer, C. A. M., Lawrence, M. G., Jeuken, A. B. M. & Velthoven, P. F. J. v. Evaluation of stratosphere–troposphere exchange and the hydroxyl radical distribution in three-dimensional global atmospheric models using observations of cosmogenic ^{14}CO . *J. Geophys. Res.* **107**, 4446, doi:10.1029/2001JD001324 (2002).
16. Schauffler, S. & Daniel, J. On the effects of stratospheric circulation changes on trace gas trends. *J. Geophys. Res.* **99**, 25747–25754 (1994).
17. Jöckel, P., Siegmund, P., Lawrence, M. G. & Brenninkmeijer, C. A. M. The detection of solar proton produced ^{14}CO . *Atmos. Chem. Phys.* **3**, 999–1005 (2003).
18. Thompson, D. W. J. & Solomon, S. Interpretation of recent southern hemisphere climate change. *Science* **296**, 895–899 (2002).
19. Forster, P. M. d. F. & Tourpali, K. Effect of tropopause height changes on the calculation of ozone trends and their radiative forcing. *J. Geophys. Res.* **106**, 12241–12251 (2001).
20. Santer, B. D. *et al.* Behaviour of tropopause height and atmospheric temperature in models, reanalyses, and observations. Part I: Decadal changes. *J. Geophys. Res.* **108**, doi:10.1029/2002JD002258 (2003).
21. Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
22. Allen, D. R., Bevilacqua, R. M., Nedoluha, G. E., Randall, C. E. & Manney, G. L. Unusual stratospheric transport and mixing during the 2002 Antarctic winter. *Geophys. Res. Lett.* **30**, 1599, doi:10.1029/2003GL017117 (2003).
23. Dlugokencky, E. J. *et al.* Changes in CH_4 and CO growth rates after the eruption of Mt Pinatubo and their link with changes in tropical tropospheric UV flux. *Geophys. Res. Lett.* **23**, 2761–2764 (1996).
24. Wooster, M. J. & Strub, N. Study of the 1997 Borneo fires: Quantitative analysis using global area coverage (GAC) satellite data. *Glob. Biogeochem. Cycles* **16**, doi:10.1029/2000GB001357 (2002).
25. Matsueda, H. & Inoue, H. Y. Aircraft measurements of trace gases between Japan and Singapore in October of 1993, 1996, and 1997. *Geophys. Res. Lett.* **26**, 2413–2416 (1999).
26. Novelli, P. C. *et al.* Reanalysis of tropospheric CO trends: Effects of the 1997–1998 wildfires. *J. Geophys. Res.* **108**, doi:10.1029/2002JD003031 (2003).
27. Lowe, D. C. *et al.* Assessment of “storage correction” required for in situ ^{14}CO production in air sample cylinders. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL014719 (2002).
28. Bergamaschi, P., Hein, R., Heimann, M. & Crutzen, P. J. Inverse modeling of the global CO cycle: 1. Inversion of CO mixing ratios. *J. Geophys. Res.* **105**, 1909–1928 (2000).
29. Jöckel, P., Lawrence, M. G. & Brenninkmeijer, C. A. M. Simulations of cosmogenic ^{14}CO using the three-dimensional atmospheric model MATCH: Effects of ^{14}C productions distribution and the solar cycle. *J. Geophys. Res.* **104**, 11733–11743 (1999).
30. Bergamaschi, P. *et al.* Transects of atmospheric CO, CH_4 , and their isotopic composition across the Pacific: Shipboard measurements and validation of inverse models. *J. Geophys. Res.* **106**, 7993–8011 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements C. Brenninkmeijer helped to establish the ^{14}CO measurement programme and R. Sparks, G. Brailsford and A. Gomez helped with the measurements and analysis used. This work was supported by the New Zealand Foundation for Research Science and Technology and Antarctica New Zealand. M.M. thanks S. Solomon for her support in completing this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.M. (Martin.Manning@noaa.gov).

In situ Os isotopes in abyssal peridotites bridge the isotopic gap between MORBs and their source mantle

Olivier Alard^{1,2,3}, Ambre Luguet^{4,5}, Norman J. Pearson¹, William L. Griffin¹, Jean-Pierre Lorand⁴, Abdelmouhcine Gannoun², Kevin W. Burton² & Suzanne Y. O'Reilly¹

Abyssal peridotites are assumed to represent the mantle residue of mid-ocean-ridge basalts (MORBs). However, the osmium isotopic compositions of abyssal peridotites and MORB do not appear to be in equilibrium^{1–8}, raising questions about the cogenetic relationship between those two reservoirs. However, the cause of this isotopic mismatch is mainly due to a drastic filtering of the data based on the possibility of osmium contamination by sea water⁵. Here we present a detailed study of magmatic sulphides (the main carrier of osmium) in abyssal peridotites and show that the $^{187}\text{Os}/^{188}\text{Os}$ ratio of these sulphides is of primary mantle origin and can reach radiogenic values suggesting equilibrium with MORB. Thus, the effect of sea water on the osmium systematics of abyssal peridotites has been overestimated and consequently there is no true osmium isotopic gap between MORBs and abyssal peridotites.

A basic tenet of isotope geochemistry is that partial melting at mantle temperature, pressure and timescales achieves complete equilibrium between melt and solid residue¹. Thus, the isotopic composition of the melt produced is identical to that of the solid residue. Abyssal peridotites (AP) are thought to represent the mantle residue after extraction of mid-ocean-ridge basalts (MORB). The observation that the Nd-isotope compositions of AP are identical to those of nearby MORB supports a direct link between AP and MORB². The unusual geochemical properties of the Re–Os system offer a new perspective on mantle processes. Unlike the other long-lived isotopic systems such as Rb–Sr or Sm–Nd, which involve incompatible lithophile elements, Os is chalcophile, behaves as a highly compatible element during melting and is retained in the mantle, whereas Re is moderately incompatible and enters the melt.

Os isotopic studies of AP and MORBs have challenged in several respects the paradigm of isotopic geochemistry. The relatively high $^{187}\text{Os}/^{188}\text{Os}$ ratio of MORB (>0.135 ; refs 3, 4) compared to 'fresh' AP ($^{187}\text{Os}/^{188}\text{Os} \leq 0.125^5$; Fig. 1) has suggested the existence of an isotopic gap between MORB and their residue. This discrepancy has fuelled a debate on the origin and significance of this Os isotopic gap and led to a variety of suggestions about the relationship between AP and MORB. Several authors^{6–8} have questioned whether AP are direct residues of recent MORB melting. It has also been suggested that veins of recycled oceanic crust or pyroxenite are preferentially melted (incongruent melting) to generate MORB^{3,9}.

However, to address the concept of an isotopic mismatch, it is crucial to consider the effect of sea water contamination on the Os systematic of AP (Fig. 1). AP show a large isotopic range

($0.116 \leq ^{187}\text{Os}/^{188}\text{Os} \leq 0.17$; refs 4, 5, 7, 10, 11), overlapping the MORB domain (Fig. 1). On the basis of the extremely radiogenic Os isotopic composition of the sea water ($^{187}\text{Os}/^{188}\text{Os} \approx 1$; ref. 12), AP with $^{187}\text{Os}/^{188}\text{Os} > 0.125$ have been dismissed as reflecting a 'secondary' contamination^{5,10}. However, sea water has extremely low Os concentrations¹² ($\sim 10^{-5}$ p.p.b.) relative to typical AP ($\sim 3\text{--}4$ p.p.b.), and large volumes of sea water are required to shift the $^{187}\text{Os}/^{188}\text{Os}$ ratio of the AP towards radiogenic values (Fig. 1).

Osmium and the other highly siderophile elements (HSE: Os, Ir, Ru, Rh, Pt, Pd, Au and Re) in the Earth's mantle are strongly concentrated in sulphides^{13–16}. Although sulphides are ubiquitous

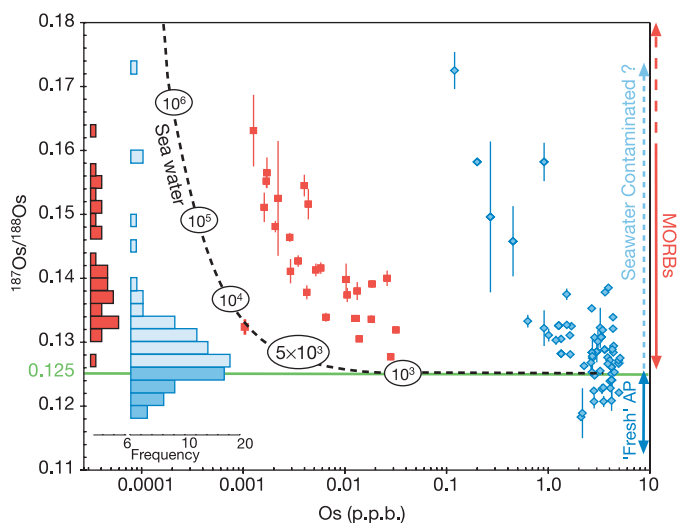


Figure 1 | Os concentration versus isotopic composition of whole-rock AP and MORB glasses. Blue diamonds, whole-rock APs^{4,5,7,10,11}; red squares, MORB glasses³. The histogram inset shows the frequency of Os isotopic composition of whole-rock APs and peridotites from the Ligurides³⁰ (in blue) and MORB glasses (in red). Error bars are 2σ . The bold green line denotes the threshold value ($= 0.125$) commonly used for 'fresh' AP⁵. AP having higher $^{187}\text{Os}/^{188}\text{Os}$ were dismissed (in light blue) as contaminated by sea water and thus not retaining a primary (mantle-derived) Os isotopic signature. The dashed curve represents mixing between typical 'fresh' AP ($\text{Os} = 3.3$ p.p.b., $^{187}\text{Os}/^{188}\text{Os} = 0.125$; ref. 5) and sea water ($\text{Os} = 1 \times 10^{-5}$ p.p.b., $^{187}\text{Os}/^{188}\text{Os} = 1$)¹²; circled numbers denote the water-rock ratio.

¹GEMOC ARC National Key Centre, School of Earth and Planetary Sciences, Macquarie University, NSW 2109, Australia. ²Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK. ³Laboratoire de Tectonophysique (cc049), Université de Montpellier II, ISTEEM CNRS UMR5560, Pl. E. Bataillon, 34095 Montpellier, France. ⁴Laboratoire 'Minéralogie, Météorites, Manteau', FREE CNRS 2456, Muséum National d'Histoire Naturelle, 61 rue Buffon, 75005 Paris, France. ⁵Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington DC 20015, USA.

in the mantle, they are rare, occur as discrete microphases and remain difficult to characterize quantitatively or qualitatively with conventional geochemical tools. Recent investigations^{7,13,16–20} have shown that the HSE inventory of the Earth's mantle is not as primitive as was first thought, but has been strongly modified during melting and melt-rock reactions due to sulphide mobility and reactivity. How much these processes have affected the Os isotopic compositions of AP remains to be fully constrained. Correlations between whole-rock HSE fractionation and Os isotopic composition of AP have been previously reported^{7,17}. This would suggest that the high $^{187}\text{Os}/^{188}\text{Os}$ ratio of some AP is mantle-derived⁷ and not due to alteration. To test these hypotheses^{5,7,10,17}, we have undertaken a detailed study of several abyssal- or oceanic-related peridotites (see Supplementary Information) using *in situ* laser ablation techniques²¹ to investigate the Os systematics of the various sulphide populations¹⁴ while preserving the petrographic and mineralogical information that is crucial for meaningful interpretations. Hereafter, for the sake of clarity, we discuss data from one representative sample (KN3-4).

KN3-4 (Kane fracture zone, Mid-Atlantic Ridge, 20–24°N) is a heavily (80%) serpentinized harzburgite ($\text{cpx}\% = 2.7\%$; $\text{Al}_2\text{O}_3 = 1.9\text{ wt}\%$, $\text{Cr}\#_{\text{spinel}} = 24.6\%$), showing limited (not pervasive) cold-seawater alteration (= sea-floor weathering, in contrast to serpentinization due to water-rock reactions occurring at higher temperatures 500–300°C). The relict primary coarse-grained assemblage consists of olivine (up to 1 cm), orthopyroxene and clinopyroxene ($\text{cpx}_1 \approx 1.3\%$) showing evidence of deformation and recrystallization, and vermicular spinel. This primary assemblage is locally overprinted by secondary aggregates of millimetre-sized clinopyroxene ($\text{cpx}_2 \approx 1.5\%$) and tiny grains of spinel (spl_2)²².

Two populations of magmatic sulphides have been recognized^{19,20,23}. Type-1 magmatic (M1) sulphides are either hosted in orthopyroxene or occur as relicts within the serpentine matrix. They consist of pentlandite $\pm$ pyrrhotite and small amounts of chalcopyrite (with pentlandite > pyrrhotite > chalcopyrite). Type-2 magmatic (M2) sulphides, spatially associated with $\text{Cpx}_2 \pm \text{Spl}_2$ clusters, consist of Ni-rich pentlandite and chalcopyrite $\pm$ bornite, with chalcopyrite + bornite $\geq 5\%$ (refs 19, 20). Abundant crisscrossed lamellae of chalcopyrite ($\leq 1\text{ }\mu\text{m}$ thick) in pentlandite contribute further to the Cu-rich composition of M2-sulphide. M2-sulphides in KN3-4 are more abundant and commonly of much larger size than M1-sulphides. All magmatic sulphides not enclosed in silicates bear the imprint of serpentinization or sea-floor weathering alteration. Typical secondary sulphide assemblages due to serpentinization are

iron-rich pentlandite $\pm$ troilite $\pm$ awaruite $\pm$ magnetite, leading to an overall metal enrichment relative to the primary assemblage. Sea-floor alteration leaches sulphur, leading to a replacement of the primary assemblage by violarite and hydroxides^{19,20,23}. Hydrothermal sulphides occur as veins of pyrite $\pm$ pyrrhotite $\pm$ marcasite.

The four whole-rock splits of KN 3-4 have $^{187}\text{Os}/^{188}\text{Os}$ between 0.1270 and 0.1309 (Supplementary Table 1). Following Snow and Reisberg⁵, this sample should not be considered as representative of the MORB residual mantle and would normally be dismissed as contaminated by sea water. However, the Os concentration is typical of mantle residues (for example, $\text{Os}_{\text{PM}} = 3.3$; ref. 24), and the Os/Ir (≈ 1.02) is also indistinguishable from that of the primitive mantle (PM), ruling out Os addition or removal via secondary processes. In contrast, the whole-rock Pd/Ir ($= 2.3$) is much higher than estimated for the PM ($\text{Pd/Ir}_{\text{PM}} = 1.2$; ref. 24). This increase is not consistent with sea-floor alteration but is symptomatic of HSE fractionation by melt percolation¹⁷. It has been shown that M1-sulphides have low Pd/Ir ratios (that is, $\text{Pd/Ir}_{\text{M1-sulphide}} < \text{Pd/Ir}_{\text{PM}}$), while M2-sulphides show high Pd/Ir ratios ($\text{Pd/Ir}_{\text{M2-sulphide}} > \text{Pd/Ir}_{\text{PM}}$; see Supplementary Table 1). The high Pd/Ir ratios of AP (including KN3-4) are due to mixtures at the hand-sample scale of the two magmatic sulphide populations^{17–20}. The $^{187}\text{Os}/^{188}\text{Os}$ ratio of magmatic sulphides in KN3-4 ranges between 0.117 and 0.167 (Supplementary Table 1) and the $^{187}\text{Re}/^{188}\text{Os}$ ratio varies from 0.07 to 1.49. Thus the Os isotopic compositions of M1- and M2-sulphides in a single AP encompass the whole range of $^{187}\text{Os}/^{188}\text{Os}$ reported for whole-rock AP worldwide, including highly radiogenic values (Fig. 1). There is no correlation between $^{187}\text{Os}/^{188}\text{Os}$ of the sulphide and the occurrence of secondary sulphide assemblages. For example, sulphide KN3-4-R, which contains about 70% hydroxides, shows a significantly less radiogenic $^{187}\text{Os}/^{188}\text{Os}$ than sulphide KN3-4-H, which is devoid of secondary phases (Supplementary Table 1). Thus, the Os isotopic composition of the magmatic sulphides are not significantly shifted during serpentinization or sea-floor alteration. Further, there is no reason why only M2-sulphides should be affected. The low Os concentration ($< 100\text{ p.p.b.}$) of hydrothermal sulphide (two to three orders of magnitude lower than M2- and M1-sulphide) prevents measurement of the $^{187}\text{Os}/^{188}\text{Os}$ ratio. However, even if they represent more than 50% by weight of the sulphide population, they would only account for $< 1\%$ of the Os budget; their effect on the whole rock $^{187}\text{Os}/^{188}\text{Os}$ is negligible.

Figure 2a shows that the whole-rock Re–Os characteristics are

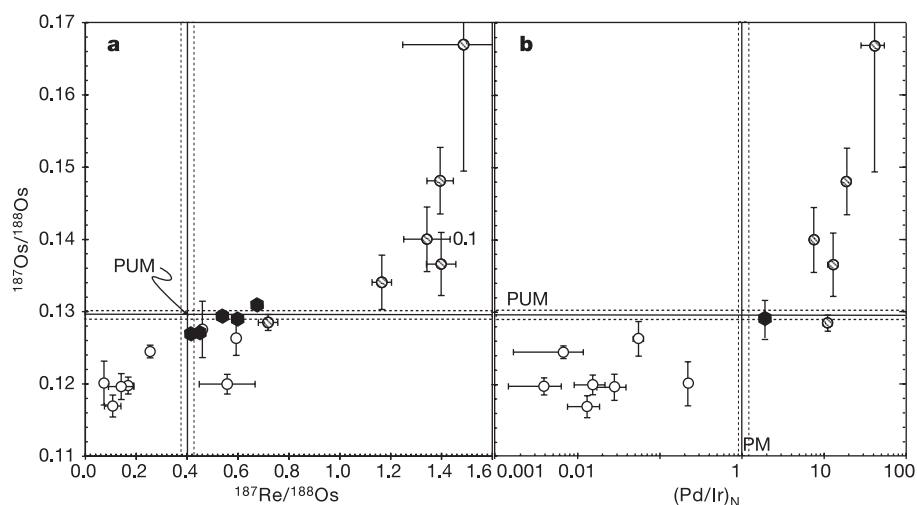


Figure 2 | $^{187}\text{Os}/^{188}\text{Os}$ versus $^{187}\text{Re}/^{188}\text{Os}$ and $(\text{Pd/Ir})_N$ for KN3-4 whole rock and sulphides. Error bars are 2 s.e.m.; PUM, primitive upper mantle value³¹; subscript N denotes PM-normalized; PM, primitive mantle value²⁴.

a, $^{187}\text{Os}/^{188}\text{Os}$ versus $^{187}\text{Re}/^{188}\text{Os}$; solid hexagonal symbols, whole-rock splits; open circles, M1-sulphides; hashed circles, coexisting M2-sulphides. **b**, $^{187}\text{Os}/^{188}\text{Os}$ versus $(\text{Pd/Ir})_N$, symbols as in **a**.

accounted for by a mixing in various proportions of unradiogenic and radiogenic sulphides, and this is illustrated at the whole-rock scale by variations between the powder splits, which reflect the heterogeneous distribution of the two populations of magmatic sulphides. Further, the relationship between the Pd/Ir and $^{187}\text{Os}/^{188}\text{Os}$ ratios (Fig. 2b) attests that the Os signature of the magmatic sulphides is of mantle origin. Indeed, as discussed above the fractionation of Pd relative to Ir is a purely magmatic process owing to the preferential partitioning of Pd into sulphide melt as Cu and Re²⁵. Thus Fig. 2b and mass balance calculations^{19,20} demonstrate unequivocally that the signature of the whole rock is fully accounted for by the mixing of the two magmatic sulphide populations both in terms of Os isotopic composition and HSE fractionation.

The mineralogy and petrographic characteristics, and the low Pd/Ir, indicate that M1-sulphides represent sulphide residual after melting^{13,16,17} as indicated by their unradiogenic Os composition ($0.1169 \leq ^{187}\text{Os}/^{188}\text{Os} \leq 0.1263$), which attests to a long-term evolution in a low Re/Os environment (that is, depleted mantle). Meanwhile, M2-sulphides show Ni-, Cu-rich bulk (high metal/sulphur) compositions indicative of sulphides derived from low-degree partial melting of mantle sulphides^{25–27}. Their systematic textural association with $\text{Cpx}_2 \pm \text{Spl}_2$ clusters and their high Pd/Ir ratio ($\gg \text{Pd/Ir}_{\text{PM}}$) demonstrates that these sulphides were precipitated during melt-rock reaction. Their radiogenic Os isotopic compositions ($^{187}\text{Os}/^{188}\text{Os} > 0.1285$) are consistent with this interpretation and suggest that those sulphides originated from a high-Re/Os reservoir (for example, pyroxenite). The crystallization of these $\text{Cpx}_2 \pm \text{Spl}_2$ patches has been ascribed to the temperature decrease at the asthenosphere–lithosphere boundary²², where the decrease in melt volume and FeO (due to the crystallization of Spl_2) will trigger the co-precipitation of M2-sulphides. Systematic studies^{18–20,22} have shown that the occurrence of M2-sulphide $\text{Cpx}_2 \pm \text{Spl}_2$ aggregates is typical of AP from slow-spreading-rate environments²² where abundant magma is trapped in the mantle⁸.

Thus, magmatic sulphides in AP show extremely variable $^{187}\text{Os}/^{188}\text{Os}$ ratios, covering the whole range so far reported for AP. In particular ‘fresh’ M2-sulphides show radiogenic Os compositions largely overlapping the MORB domain (that is, > 0.130). The radiogenic values of the M2-sulphides are not inherited or affected by sea water. Seawater contamination of the percolating melt or of the peridotite residue cannot produce the Pd/Ir– $^{187}\text{Os}/^{188}\text{Os}$ correlation that is a signature of a purely magmatic process. The whole-rock Os isotopic compositions of the AP—as with their HSE characteristics—reflect mixing between several magmatic sulphide populations showing contrasting mineralogical and geochemical characteristics. Sulphide/whole-rock/HSE/Re–Os systematics similar to those in KN3-4 have been found in samples from the southwest Indian ridges and the Ligurides ophiolites, representing the Jurassic Liguro-piemontaise ocean floor (Supplementary Table 1) and thus support the worldwide occurrence of a radiogenic mantle sulphide population within the upper oceanic mantle.

There is thus no need systematically to invoke seawater contamination to account for the radiogenic Os composition of some AP. The threshold value of 0.125, previously used to differentiate ‘fresh’ from ‘altered’ AP, is not justified and should not be used as such without supporting petrological and geochemical information (although we note that it was not the intention of the original paper⁵ to establish this value as a cut-off). This is not to deny the dramatic effect of sea water, during both serpentinization and sea-floor alteration, on the mineralogy and geochemistry of AP. Rather, its effect on the Os isotopic systematic of the AP has been greatly overestimated, biasing our estimate of the Os composition of AP worldwide, and hence the depleted MORB mantle. Consequently, the so-called ‘Os isotopic gap’ most probably does not exist and the ‘Os paradox’ between MORB and AP is resolved. Although very radiogenic compositions for both MORB and AP may well result from seawater alteration²⁸, for MORBs, if necessary²⁹, those compositions might be explained by

the preferential mobilization (incongruent melting, melt-percolation reaction) of the M2 sulphides²⁵. Thus we conclude that MORBs are globally in isotopic equilibrium with AP, their assumed mantle residue.

Received 17 April; accepted 7 June 2005.

- Hofmann, A. W. in *The Mantle* (ed. Carlson, R. W.) 61–101 (Elsevier, Oxford, 2004).
- Snow, J., Hart, S. R. & Dick, H. B. J. Nd and Sr isotopic evidence for a link between mid-ocean ridge basalts and abyssal peridotites. *Nature* **371**, 57–60 (1994).
- Schiano, P., Birck, J.-L. & Allègre, C. J. Osmium-strontium-neodymium-lead isotopic covariations in mid-ocean ridge basalt glasses and the heterogeneity of the upper mantle. *Earth Planet. Sci. Lett.* **150**, 363–379 (1997).
- Roy-Barman, M. & Allègre, C. J. $^{187}\text{Os}/^{186}\text{Os}$ ratios of mid-ocean ridge basalts and abyssal peridotites. *Geochim. Cosmochim. Acta* **58**, 5043–5054 (1994).
- Snow, J. E. & Reisberg, L. Os isotope systematics of the MORB mantle: results from altered abyssal peridotites. *Earth Planet. Sci. Lett.* **133**, 411–421 (1995).
- Esperança, S., et al. *Ninth V.M. Goldschmidt Conf.* Lunar and Planetary Institute, Contrib No. 971, abstract No. 7389, (Houston, 1999).
- Brandon, A. D., Snow, J. E., Walker, R. J., Morgan, J. W. & Mock, T. D. ^{190}Pt – ^{186}Os and ^{187}Re – ^{187}Os systematics of abyssal peridotites. *Earth Planet. Sci. Lett.* **177**, 319–335 (2000).
- Seyler, M., Lorand, J.-P., Toplis, M. J. & Godard, G. Asthenospheric metasomatism beneath the mid-oceanridge: evidence from depleted abyssal peridotites. *Geology* **32**, 301–304 (2004).
- Hirschmann, M. M. & Stolper, E. M. A possible role for garnet pyroxenite in the origin of the ‘garnet signature’ in MORB. *Contrib. Mineral. Petrol.* **124**, 185–208 (1996).
- Standish, J. J., Hart, S. R., Blusztajn, J., Dick, H. B. J. & Lee, K. L. Abyssal peridotite osmium isotopic composition from Cr-spinel. *Geochim. Geophys. Geosyst.* **3**, doi:10.1029/2001GC000161 (2002).
- Martin, C. E. Osmium isotopic characteristics of mantle-derived rocks. *Geochim. Cosmochim. Acta* **55**, 1421–1434 (1991).
- Levasseur, S., Birck, J.-L. & Allegre, C. J. Direct measurement of femtomoles of osmium and the $^{187}\text{Os}/^{186}\text{Os}$ ratio in seawater. *Science* **282**, 272–274 (1998).
- Alard, O., Griffin, W. L., Lorand, J.-P., Jackson, S. E. & O'Reilly, S. Y. Non-chondritic distribution of the highly siderophile elements in mantle sulphides. *Nature* **407**, 891–894 (2000).
- Alard, O., Griffin, W. L., Pearson, N. J., Lorand, J.-P. & O'Reilly, S. Y. New insights into the Re–Os systematics of sub-continental lithospheric mantle from in situ analysis of sulphides. *Earth Planet. Sci. Lett.* **203**, 651–663 (2002).
- Burton, K. W., Schiano, P., Birck, J.-L. & Allègre, C. J. Osmium isotope disequilibrium between mantle minerals in a spinel-lherzolite. *Earth Planet. Sci. Lett.* **172**, 311–322 (1999).
- Lorand, J.-P. & Alard, O. Platinum-group element abundances in the upper mantle: new constraints from in situ and whole-rock analyses of Massif Central xenoliths (France). *Geochim. Cosmochim. Acta* **65**, 2789–2806 (2001).
- Rehkämper, M. et al. Non-chondritic platinum-group element ratios in oceanic mantle lithosphere: Petrogenetic signature of melt percolation? *Earth Planet. Sci. Lett.* **172**, 65–81 (1999).
- Luguet, A., Lorand, J.-P., Alard, O. & Cottin, J.-Y. A multi-technique study of platinum-group elements systematic in some ligurian ophiolitic peridotites. *Chem. Geol.* **208**, 175–194 (2004).
- Luguet, A. et al. Laser-ablation microprobe (LAM)-ICPMS unravels the highly siderophile element geochemistry of the oceanic mantle. *Earth Planet. Sci. Lett.* **189**, 285–294 (2001).
- Luguet, A., Lorand, J.-P. & Seyler, M. Sulfide petrology and highly siderophile element geochemistry of abyssal peridotites: a coupled study of samples from the Kane Fracture Zone (45°W 23°20N, MARK area, Atlantic Ocean). *Geochim. Cosmochim. Acta* **67**, 1553–1570 (2003).
- Pearson, N. J., Alard, O., Griffin, W. L., Jackson, S. E. & O'Reilly, S. Y. In situ measurement of Re–Os isotopes in mantle sulfides by laser ablation multicollector-inductively coupled plasma mass spectrometry: analytical methods and preliminary results. *Geochim. Cosmochim. Acta* **66**, 1037–1050 (2002).
- Seyler, M., Toplis, M. J., Lorand, J.-P., Luguet, A. & Cannat, M. Clinopyroxene microtextures reveal incompletely extracted melts in abyssal peridotites. *Geology* **29**, 155–158 (2001).
- Luguet, A. & Lorand, J.-P. Fe–Ni–Cu sulphide of abyssal peridotites from the MARK area (Mid-Atlantic Ridge, 20–24°N). *C.R. Acad. Sci. IIA* **329**, 637–644 (1999).
- McDonough, W. F. & Sun, S.-S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Bockrath, C., Ballhaus, C. G. & Holzheid, A. Fractionation of the platinum-group elements during mantle melting. *Science* **305**, 1951–1953 (2004).
- Ballhaus, C., Tredoux, M. & Spaeth, A. Phase relation in the Fe–Ni–Cu–PGE–S system at magmatic temperature and application to massive sulfide ores of Sudbury Igneous Complex. *J. Petrol.* **42**, 1991–2026 (2001).

27. Craig, J. R. & Kullerud, G. in *Magmatic Ore Deposits* (ed. Wilson, H. D. B.) 343–358 (Econ. Geol. Monogr., Geological Survey of Canada, Ontario, 1989).
28. Gannoun, A. *et al.* Osmium isotope heterogeneity in the constituent phases of mid-ocean ridge basalts. *Science* **303**, 70–72 (2004).
29. Gannoun, A., Burton, K. W., Alard, O., Parkinson, I. J. & Thomas, L. E. Assessing the scale of osmium isotope heterogeneity in mid-ocean ridge basalts. *Geochim. Cosmochim. Acta* **68** (suppl. 1), 703 (2004).
30. Snow, J. E., Schmidt, G. & Rampone, E. Os isotopes and highly siderophile elements (HSE) in the Ligurian ophiolites, Italy. *Earth Planet. Sci. Lett.* **175**, 119–132 (2000).
31. Meisel, T., Walker, R. J. & Morgan, J. W. The osmium isotopic composition of the Earth's upper mantle. *Nature* **383**, 517–520 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements O.A. thanks the Royal Society for financial support. Discussions with M. Godard, J. L. Bodinier, M. Gounelle and L. Reisberg helped to improve the manuscript. This is an ARC National Key Centre for Geochemical Evolution and Metallogeny of Continents (GEMOC) publication.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.A. (o.alard@open.ac.uk).

Earthquake rupture dynamics frozen in exhumed ancient faults

Giulio Di Toro^{1*}, Stefan Nielsen^{2*} & Giorgio Pennacchioni¹

Most of our knowledge about co-seismic rupture propagation is derived from inversion and interpretation of strong-ground-motion seismograms^{1–3}, laboratory experiments on rock^{4,5} and rock-analogue material⁶, or inferred from theoretical and numerical elastodynamic models^{7–9}. However, additional information on dynamic rupture processes can be provided by direct observation of faults exhumed at the Earth's surface¹⁰. Pseudotachylytes (solidified friction-induced melts^{11,12}) are the most certain fault-rock indicator of seismicity on ancient faults¹³. Here we show how the asymmetry in distribution and the orientation of pseudotachylyte-filled secondary fractures around an exhumed fault can be used to reconstruct the earthquake rupture directivity, rupture velocity and fracture energy, by comparison with the theoretical dynamic stress field computed around propagating fractures. In particular, the studied natural network of pseudotachylytes is consistent with a dominant propagation direction during repeated seismic events and subsonic rupture propagation close to the Rayleigh wave velocity.

The studied natural example, the Gole Larghe fault, is a dextral strike-slip fault zone crosscutting the Tertiary (42–30-Myr-old) Adamello intrusion¹⁴. The Adamello consists of tonalite, a granitoid rock that serves as a representative rock of the Earth's upper continental crust. The Gole Larghe fault branches off a bend of the Tonale fault, which is a segment of the major tectonic lineament of the Alps¹⁵ (Fig. 1). The exposed sections of Gole Larghe fault underwent deformation at 9–11 km depth and 250–300 °C ambient conditions about 30 Myr ago (refs 14, 16). The fault zone was not reactivated (except very locally) or tilted during subsequent uplift and exhumation. The palaeoseismicity of the Gole Larghe fault is attested by the presence of pseudotachylytes.

In the upper Val di Genova, the fault zone of the Gole Larghe is about 550 m thick and spectacularly exposed in glacier-polished outcrops (star in Fig. 1). The fault zone consists of an array of about 200 major sub-parallel faults that are spaced 2–6 m apart, strike about N100° and dip about 50° towards N190°. Fault rocks include cataclasites (produced by brittle deformation without melting) overprinted by pseudotachylytes¹⁴. Slip along major faults is accompanied by the development of a damage zone network of minor faults and fractures in the host tonalites. Kinematic indicators show dextral strike-slip movement during both cataclasite-related and pseudotachylyte-related faulting. Faults nucleated on pre-existing segmented joints, which form a pervasive set in the Adamello¹⁴. High geometric connectivity between originally segmented joints was achieved during the precursory cataclastic faulting, by abrasion of tonalite indentations at *en-echelon* step-overs¹⁴. Cataclastic deformation was the prelude to seismic fracture propagation by developing roughly 10-km-scale relatively weak sub-parallel and continuous faults⁴. Individual fault displacement ranges from a few centimetres to 30 m, leading to a cumulative offset of about 1 km across the whole

fault zone¹⁴. Fault segments carrying only pseudotachylytes (that is, not associated with cataclastic precursor) have displacements of less than 1.5 m, which are typical of seismic fault ruptures of about 10 km in length¹⁷. Field and microstructural data indicate that pseudotachylyte is produced at the final stage of fault slip¹⁴. Evidence of multiple generations of pseudotachylytes is rare along the Gole Larghe fault¹⁴; pseudotachylytes within each individual fault segment are therefore the result of a single seismic rupture.

Many pseudotachylyte veins were injected into the host tonalite from pseudotachylyte-bearing fault segments; an example of a fault segment with injection veins is shown in Fig. 2. We measured the orientation of 624 injection veins that branch off 28 different fault segments, in exposures sub-parallel to the fault slip direction. Linear fault segments that were at least 2–3 m away from the closest segment were selected for the measures, to avoid potential perturbations due to fault irregularity¹⁸ and interference with neighbouring faults. The cumulative data from all measures (Fig. 3) reveals two dominant orientations of injection veins, at about 30–210° (referred to as set 1) and 90–270° (set 2) with respect to the fault trace. Both vein sets are asymmetrically distributed with respect to the fault trace with distinct dominance (67.7%) of the veins intruded into the southern bounding block. The veins of set 1 (mostly less than 2 mm thick and less than 50 cm long) intruded pre-existing minor cataclastic faults

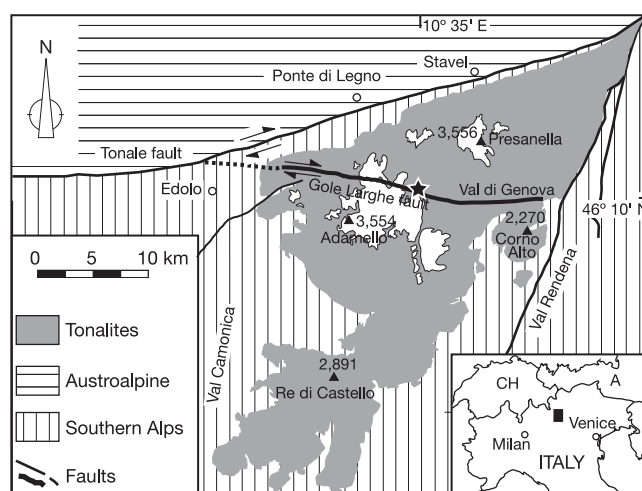


Figure 1 | Geological sketch map of the Adamello intrusion. The Tonale fault marks the boundary between two main domains of the Alpine chain, namely the south-verging thrust belt of the southern Alps and north-verging Austroalpine units. The Gole Larghe fault (thick line) intersects the Tonale fault north of Edolo close to the bend of the Tonale fault from E–W to ENE–WSW. The star indicates the location of the outcrop under study.

¹Dipartimento di Geologia, Paleontologia e Geofisica - Università di Padova, 35137 Padova, Italy. ²Istituto Nazionale di Geofisica e Vulcanologia (INGV), 00143 Roma, Italy.

*These authors contributed equally to the work.

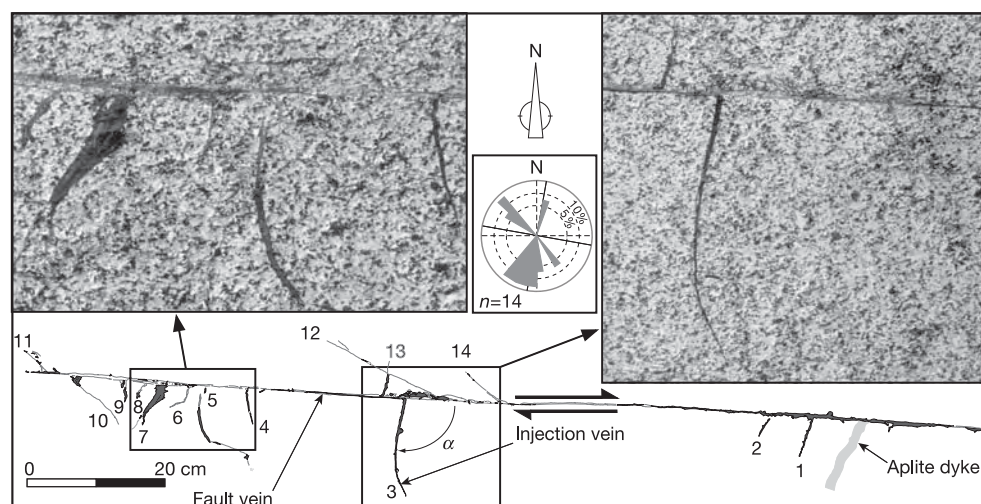


Figure 2 | Pseudotachylyte-bearing fault. Pseudotachylytes are easily distinguished in the field by their dark colour (see enlargements in the photos). The orientations of pseudotachylyte injection veins (numbered 1 to

14) are shown in the area-weighted rose diagram in the inset. The injection vein orientation α was measured clockwise from the east side of fault. The sense of the shear is dextral.

oriented at about $15\text{--}25^\circ$ to the main fault surface. The veins of set 2 (commonly more than 2 mm thick and less than 30–40 cm long) intruded new fractures that have no counterpart in the array of discontinuities of the cataclastic damage zone¹⁴. Set 2 is therefore interpreted as a group of tension fractures that developed during propagation of the co-seismic rupture. The veins that intruded the southern fault-bounding block are dominant in almost all the measured fault segments (see, for example, Fig. 2), and in the cumulative data set (Fig. 3). Of the 28 fault segments studied, 25 faults have more fractures towards the south. In particular, pseudotachylyte veins injecting into the southern block are more than 70% on 17 fault segments, 60–70% on three segments and 50–60% on five segments. When considering only those fractures of pure co-seismic origin (set 2), the dominant southern proportion is even more pronounced. (The data set of pseudotachylyte vein orientation and location of the fault segments studied is given in Supplementary Information.)

The observed asymmetry of the pseudotachylyte veins with respect to the fault segments can be explained by the asymmetry in the stress field around the tip of a propagating Mode II (in-plane shear) fracture; such a fracture displays transient compression and tension fields on the opposite sides of the rupture plane^{7–9}. As the strength of rocks in tension is only one-tenth of that in compression, most fractures and damage are anticipated to develop on the fault side under tension^{19,20}.

The relationship between the secondary fractures and the transient tension field is investigated here by computing the dynamic stress pattern associated to a propagating rupture. For simplicity we adopt a model of single, planar rupture, propagating at fixed velocity. Given a fixed pre-stress direction, the dynamic stress field around fracture strongly depends on the ratio of rupture propagation velocity, v , to the shear wave velocity, v_s (refs 7–9). We considered three cases: first, slow rupture velocities ($v = 0.6v_s$), indicative of large dissipation in the fracture process; second, high rupture velocities ($v = 0.9v_s$), about 98% of the Rayleigh wave velocity, an important asymptotic limit in the subsonic regime²¹; and third, supershear rupture velocities ($v = 2^{1/2}v_s$), which correspond to a stable, supersonic fracture solution observed in laboratory experiments⁶ and predicted theoretically^{22,23}.

The solution proposed previously⁹, valid for subsonic rupture propagation, allows the stress field around a two-dimensional in-plane steady-state pulse to be computed by explicitly accounting for a slip-weakening zone of arbitrary length at the fracture tip. For

supershear cases, we adopted a classical finite-difference time domain scheme used for solving dynamic faulting problems³. Transient fluctuations in pore-fluid pressure were not considered, as microstructures and geochemistry suggest that pseudotachylytes developed under fluid-deficient conditions^{14,16}. The static stress tensor at the time of seismicity has been estimated¹⁶ on the basis of deformation depth (about 10 km) and fault mechanics. This yielded, in the horizontal plane, normal stresses of -112 and -208 MPa (perpendicular and parallel to the fault, respectively) and a shear stress of 83 MPa on the fault. The constitutive properties used in the model were those of the tonalitic host rock²⁴: a Poisson ratio of 0.25, a shear modulus of 26 GPa, a fracture toughness of $2 \text{ MPa m}^{1/2}$ and rock density of $2,700 \text{ kg m}^{-3}$. We considered two limiting values of the decrease in dynamic stress: 5.6 and 42 MPa (ref. 16). For any given combination of decrease in dynamic stress and fracture velocity, the size of the slip pulse was designed to obtain a seismic slip close to

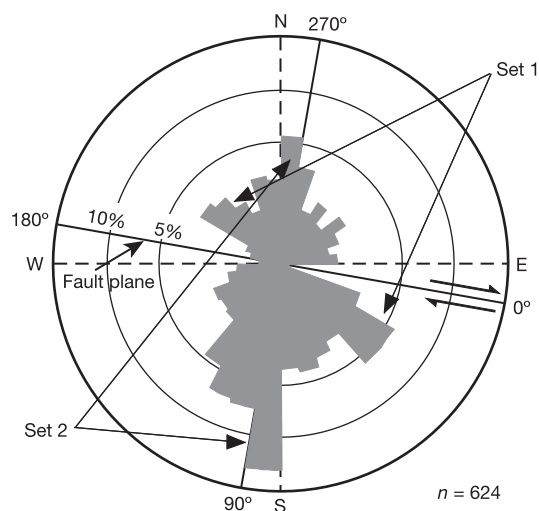


Figure 3 | Orientation of pseudotachylyte injection veins along 28 different faults, in an area-weighted rose diagram. The angular reference frame is relative to the fault plane orientation (the $0^\circ\text{--}180^\circ$ direction) and measurements are relative to α as defined in Fig. 2. Set 1 and set 2 veins are indicated by arrows. The fault-plane strike is close to the east–west direction (dashed lines indicate the geographic reference frame). The sense of the shear is dextral for all faults. Most veins intrude the southern wall rock.

1.5 m (the maximum measured displacement in pseudotachylite-bearing faults without evidence of a precursor cataclase). The resulting fracture pulses are a few hundred metres to 1 km long. A linear slip-weakening region was adjusted in each model to obtain values of the peak shear stress in agreement with the shear strength of granitoid rocks (100–200 MPa) under the assumed normal stress.

All the elastodynamic solutions yield a tension region located in the southern block close to the propagating fracture tip for ruptures propagating from west to east (Fig. 4). In spite of the large lithostatic load, the computed dilation is large enough to induce Mode I cracks for subsonic propagation velocities. When rupture approaches the Rayleigh wave velocity, constructive interference is maximized and

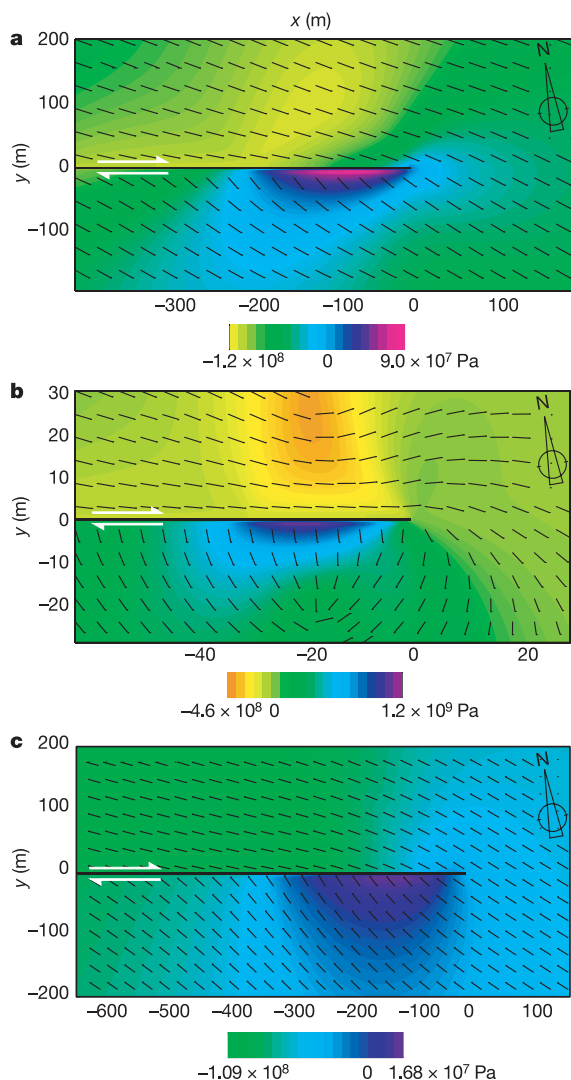


Figure 4 | Tensile stress field during rupture propagation in the vicinity of a fracture tip for three different rupture velocities. The fracture tip is shown as a black line. The velocities were $v = 0.6v_s$ (a), $v = 0.9v_s$ (b) and $v = 2^{1/2}v_s$ (c). The three diagrams are zoom boxes, around the fault tip, of the original model (note the different magnifications). The model assumes pre-stress and constitutive parameters based on the Gole Larghe fault at the time of seismicity. Fracture propagation is from west to east on a vertical dextral strike-slip fault. Colour shows stress magnitude (in Pa), from orange (negative, compression) to blue (positive, tension); thin black segments indicate the local direction of planes of maximum tension. In all cases, a tensional domain of variable amplitude and extension (40–200 m) is present on the southern side of the rupture. In **b** the maximum tension planes are oriented nearly orthogonal to the fault plane, which is consistent with orientation of most pseudotachylite-filled tension cracks observed in the southern bounding block along the Gole Larghe fault (Fig. 3).

the stress perturbation, contracting around the crack tip, is magnified (a process known elsewhere as the Lorentz contraction²⁵). In natural faults the excessive peak tension values obtained in the models are hindered by the instantaneous development of tension cracks. The examples shown in Fig. 4 correspond to a stress drop of 42 MPa. Reducing the stress drop does not substantially alter the results in subsonic regimes. Only for a combination of supersonic rupture and small stress drop (5.6 MPa) is the transient tension insufficient to overcome the initial compression. The direction of major tension planes depends critically on the fracture velocity. Solutions for fracture propagating close to the Rayleigh wave velocity ($v = 0.9v_s$; Fig. 4b) yield preferred orientations of tension cracks perpendicular to the fault, which is consistent with the observations along the Gole Larghe faults (Fig. 3).

A few veins also intrude the northern fault block; the overall pseudotachylite pattern therefore cannot be explained solely by direction and velocity propagation of seismic fracture ($v = 0.9v_s$), unless additional complexity is invoked in the model. Indeed, non-planar geometry, strong inhomogeneities and the interaction of neighbouring faults induce a degree of complexity that is not accounted for in our model. Because tonalite (and cataclase) undergo a 17% expansion during melting²⁶, co-seismic hydrofracturing due to melt overpressure might also contribute to the generation of pseudotachylite-bearing fractures²⁷. In simulations at supershear fracture velocity (Fig. 4c), the principal compressive direction reaches maximum orientations of 60° to the fault plane; this result suggests that supersonic ruptures were infrequent during earthquakes along the Gole Larghe fault.

Those fracture models that scale with a slip of about 1.5 m, in the studied velocity range, indicate that G , the energy dissipated during fracture propagation, is between 8.0 and 67 MJ m⁻². These G values are within the high range of the fracture energy estimated from seismic data^{7,9,28} and gouge texture²⁹.

Thus, on the basis of our field observations we designed a dynamic model to explain the preferential asymmetric occurrence of pseudotachylite injection veins orthogonal to a leading fault plane. The natural pseudotachylite pattern is compatible with seismic ruptures that usually propagated eastwards at velocities approaching the Rayleigh wave velocity. The rupture propagation occurred along weak discontinuities, inducing melting and hydrofracturing of the bounding blocks. The dominant eastward propagation of palaeoearthquakes along the Gole Larghe fault is deduced from the asymmetric arrangement of pseudotachylite veins along most fault segments. Why would most ruptures along the Gole Larghe fault propagate eastwards? This challenging question is open for discussion. Because the instrumented seismological record of repeated large earthquakes along a given fault is rather small, there are no clear indications of repeated earthquake directivity. Although theoretical analyses predict that a material contrast across the fault could favour consistent rupture directivity³⁰, this model is not applicable here because the earthquakes along the Gole Larghe fault occurred within a homogeneous tonalite. We speculate that most earthquakes originated along the Tonale fault (Fig. 1), where different basement and cover units are in contact (resulting in a rheological contrast across the fault), or simply along a weaker fault segment, and propagated eastwards along the Gole Larghe fault. The present study shows that networks of co-seismically generated fractures in exhumed faults are potential gauges for constraints on dynamic parameters (rupture directivity, G and v) and represent a fundamental contribution of field geology to understanding earthquake mechanics.

Received 15 December 2004; accepted 9 June 2005.

1. Archuleta, R. J. Faulting model for the 1979 Imperial Valley earthquake. *J. Geophys. Res.* **89**, 4559–4585 (1984).
2. Heaton, T. H. Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Phys. Earth Planet. Inter.* **64**, 1–20 (1990).

3. Peyrat, S., Olsen, K. & Madariaga, R. Dynamic modeling of the 1992 Landers earthquake. *J. Geophys. Res.* **106**, 26467–26482 (2001).
4. Laqueche, H., Rousseau, A. & Valentin, G. Crack propagation under Mode I and II loading in slate schist. *Int. J. Rock Mech.* **23**, 347–354 (1986).
5. Cox, S. J. D. & Scholz, C. H. An experimental study of shear fracture in rocks: mechanical observations. *J. Geophys. Res.* **93**, 3307–3320 (1988).
6. Rosakis, A. J., Samudrala, O. & Coker, D. Cracks faster than shear wave speed. *Science* **284**, 1337–1340 (2000).
7. Poliakov, A. N. B., Dmowska, R. & Rice, J. R. Dynamic shear rupture interactions with fault bends and off-axis secondary faulting. *J. Geophys. Res.* **107**, doi:10.1029/2001JB000572 (2002).
8. Kame, N., Dmowska, R. & Rice, J. R. Effects of pre-stress and rupture velocity on dynamic fault branching. *J. Geophys. Res.* **108**, doi:10.1029/2002JB002189 (2003).
9. Rice, J. R., Sammis, C. G. & Parsons, R. Off-fault secondary failure induced by a dynamic slip-pulse. *Bull. Seismol. Soc. Am.* **95**, 109–134 (2005).
10. Allen, J. L. A multi-kilometer pseudotachylite system as an exhumed record of earthquake rupture geometry at hypocentral depths (Colorado, USA). *Tectonophysics* **402**, 37–54 (2005).
11. McKenzie, D. & Brune, J. N. Melting on fault planes during large earthquakes. *Geophys. J. R. Astron. Soc.* **29**, 65–78 (1972).
12. Sibson, R. H. Generation of pseudotachylite by ancient seismic faulting. *Geophys. J. R. Astron. Soc.* **43**, 775–794 (1975).
13. Snoke, A. W., Tullis, J. & Todd, V. *Fault-related Rocks; A Photographic Atlas* (Princeton Univ. Press, Princeton, New Jersey, 1998).
14. Di Toro, G. & Pennacchioni, G. Fault plane processes and mesoscopic structure of a strong-type seismogenic fault in tonalites (Adamello batholith, Southern Alps). *Tectonophysics* **402**, 54–79 (2005).
15. Schmid, S. M., Aebli, H. R., Heller, F. & Zingg, A. in *Alpine Tectonics* (eds Coward, M. P., Dietrich, D. & Park, R. G.) 153–171 (Geological Society Special Publication 45, London, 1989).
16. Di Toro, G., Pennacchioni, G. & Teza, G. Can pseudotachylites be used to infer earthquake source parameters? An example of limitations in the study of exhumed faults. *Tectonophysics* **402**, 3–20 (2005).
17. Sibson, R. H. Earthquake faulting as a structural process. *J. Struct. Geol.* **11**, 1–14 (1989).
18. Chester, F. M. & Chester, J. S. Stress and deformation along wavy frictional faults. *J. Geophys. Res.* **105**, 23421–23430 (2000).
19. Andrews, J. D. Rupture dynamics with energy loss outside the slip zone. *J. Geophys. Res.* **110**, 10.1029/2004JB003191 (2005).
20. Dalguer, L. A., Irikura, K. & Riera, J. Simulation of tensile crack generation by three-dimensional dynamic shear rupture propagation during an earthquake. *J. Geophys. Res.* **108**, 10.1029/2001JB001738 (2003).
21. Burridge, R., Conn, G. & Freund, L. B. The stability of a rapid mode II shear crack with finite cohesive traction. *J. Geophys. Res.* **83**, 2210–2222 (1979).
22. Broberg, K. B. The near tip field at high crack velocities. *Int. J. Fracture* **39**, 1–13 (1989).
23. Freund, L. B. The mechanics of shear crack propagation. *J. Geophys. Res.* **84**, 2199–2209 (1979).
24. Rzhetsky, Y. & Novik, G. *The Physics of Rocks* (MIR, Moscow, 1971).
25. Ida, Y. Stress concentration and unsteady propagation of longitudinal shear cracks. *J. Geophys. Res.* **78**, 3418–3429 (1973).
26. Spera, F. K. in *Encyclopedia of Volcanoes* (ed. Sigurdsson, H.) 171–190 (Academic, San Diego, 2000).
27. Swanson, M. T. Fault structure, wear mechanisms and rupture processes in pseudotachylite generation. *Tectonophysics* **204**, 223–242 (1992).
28. Li, V. C. in *Fracture Mechanics of Rocks* 2nd edn (ed. Atkinson, B. K.) 351–428 (Academic, London, 1989).
29. Wilson, B., Dewers, T. A., Reches, Z. & Brune, J. Texture and energetics of gouge powder from earthquake rupture zones. *Nature* **434**, 749–752 (2005).
30. Ben-Zion, Y. & Andrews, J. D. Properties and implications of dynamic rupture along a material interface. *Bull. Seismol. Soc. Am.* **88**, 1085–1094 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by INGV (Italy), MIUR (Italy) and the CRdC AMRA Project. We thank M. Bouchon, M. Cocco, J. Rice, T. Tullis and J. Allen for encouragement and comments; and Z. Reches for a constructive review.

Author Contributions G.D.T. and S.N. collected the field data; S.N. produced the mathematical models; G.D.T., S.N. and G.P. wrote the manuscript. G.D.T. and S.N. contributed equally to the work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.N. (snielsen@na.infn.it).

Silurian brachiopods with soft-tissue preservation

Mark D. Sutton¹, Derek E. G. Briggs², David J. Siveter³ & Derek J. Siveter^{4,5}

'Articulated' rhynchonelliformean¹ brachiopods are abundant shelly fossils, but the direct fossil record of their soft parts was hitherto confined to a single pyritized trace possibly representing a lophophore². Anatomical knowledge of extinct rhynchonelliformeans relies heavily on analogies to extant species; these analogies are untested for stem-group clades. The Silurian Herefordshire (UK) Konservat-Lagerstätte³ (about 425 Myr BP) yields exceptionally preserved three-dimensional fossils that provide unrivalled insights into the palaeobiology of a variety of invertebrates^{4–9}. The fossils are preserved as calcitic void in-fills in carbonate concretions within a volcanoclastic horizon¹⁰, and are reconstructed digitally¹¹. Here we describe a stem-group rhynchonelliformean specimen from this deposit; it most probably belongs in the order Orthida. A robust ridged pedicle with distal rootlets is preserved, together with a lophophore and other soft-tissue structures. The pedicle morphology is novel, urging caution in inferring stem-group rhynchonelliformean anatomy from that of crown-group species. Smaller brachiopods are attached to the specimen; these include a probable atrypide, with pedicle and marginal setae preserved.

Phylum Brachiopoda
Subphylum Rhynchonelliformea
Class Rhynchonellata (?)
Order Orthida (?)

Bethia serraticulma Sutton, Briggs,
Siveter and Siveter gen. et sp. nov.

Etymology. *Bethia* for Bethia Thomas; *serraticulma* from Latin *serratus* (serrated) plus *culmen* (ridge of a roof; also a stalk), alluding to high growth-lamellae and pedicle-ridges.

Holotype. Oxford University Museum of Natural History, OUM C.29586. Figure 1a–c, e–i, k, reconstructed in three dimensions. No other material is known.

Stratigraphy and locality. Wenlock Series, Silurian; Herefordshire, England.

Diagnosis for genus and species. Ventral valve strongly convex with wide sulcus, dorsal valve weakly concave. Commissural outline slightly transverse; posterolateral margins weakly invaginated; alae short, pointed; anterior and lateral margins smoothly rounded. Ventral interarea high, concave, apsacline. In juvenile, pedicle emerges subapically through a large open delthyrium, flanked dorsolaterally by incipient deltidial plates. Pseudodeltidium absent. Dorsal interarea hypercline, notothyrium closed. Dorsal valve with saccate vascular system. Thin-shelled; external ornament of high growth-lamellae; radial component absent.

Description. The shell is concavoconvex, 6.1 mm in both maximum width and length, with the widest point 68% of valve length from the umbo (Fig. 1a–c). Hinge line strophic, 75% of maximum shell width. The outline of the commissure is rounded except posterolaterally, where the cardinal extremities are produced into pointed alae (Fig. 1a, c). In both valves the protegulum (Fig. 1c, f) is subcircular and about 0.9 mm in diameter. Ornament consists of high concentric

growth-lamellae (Fig. 1a–c, f, h, k) about 0.3–0.4 mm apart medially; these are higher and have thickened tops in the ventral valve (Fig. 1f). Radial ornament is absent. Valves 60 µm or less in thickness away from muscle insertions, etc. (Fig. 1k).

The ventral valve is 2.5 mm in maximum depth, at about 35% of valve length from the umbo (Fig. 1b). The umbo faces posteriorly, overhanging a well-defined triangular apsacline interarea (Fig. 1f), 4.8 mm wide and 1.4 mm high. The apical angle is 94°. The delthyrium is large and open, occupying about 35% of the interarea width (Fig. 1f). It is bounded dorsally by two spine-like projections of the ventral valve, converging medially (Fig. 1b, c, f), interpreted as incipient deltidial plates. The valve is evenly convex away from the umbo and the alae, except for a shallow sulcus representing about 25% of the valve width (Fig. 1a, f, h, k), originating near the protegulum. Internally, the valve has subanterolateral thickened regions (Fig. 1g; clearest on right), interpreted as distal tips of sublinear structures (?diductor muscle platforms) diverging at about 65° from the umbo. Further muscle-field detail is obscured by soft tissues.

The dorsal valve is evenly concave in transverse section, with a maximum depth of 1.1 mm about 40% of valve length from the posterior. The interarea (Fig. 1c, f) is hypercline and short, with an apical angle of about 165°. The notothyrium is closed; discrete chilidial plates are not apparent. Muscle fields are not preserved.

The robust pedicle (Fig. 1a–c, i) is cylindrical and straight, 3.0 mm long and 1.0 mm in diameter; it projects posteriorly from the delthyrium. It bears an irregular series of subtransverse ridges up to 0.2 mm high that cross-cut in a disordered fashion and are lower and more closely spaced distally. The pedicle terminates abruptly in a flattened disc, from which about 25 'rootlets' arise submarginally, each about 0.05 mm in diameter and up to 1.3 mm long (Fig. 1i). The disc is in contact (off-centre) with a cylindrical piece of skeletal debris around which most of the rootlets are wrapped. This object is several centimetres long (not reconstructed in entirety) and about 0.7 mm in diameter.

The mantle is preserved only in the dorsal valve (Fig. 1e); local thickenings are interpreted as canals. The vascular pattern is ?saccate, but unclear in detail (see Supplementary Movie 2); lateral perimarginal canals, which curve inward and backward medially, are interpreted as *vascula myaria*, and *vascula media* may also be present medially. The mantle is dissociated from the shell anteriorly; this is interpreted as a taphonomic effect.

Internal soft tissues are poorly preserved posteriorly; the undifferentiated internal mass (white in Fig. 1e) is in part an artefact of incomplete sediment penetration. Two lateral lobes project from this mass, elliptical in transverse section with a ventromedial–dorsolateral long axis. These are interpreted as visceral material incorporating the lophophore marginally, evident as a broad ventrolateral ridge on one lobe (Fig. 1e). Each lophophore arm bears a single row of fine linear projections (Fig. 1e, g), interpreted as tentacles. There are about 130–150 tentacles, preserved impermissibly along

¹Department of Earth Science and Engineering, Imperial College, London SW7 2AZ, UK. ²Department of Geology and Geophysics, Yale University, PO Box 208109, New Haven, Connecticut 06520-8109, USA. ³Department of Geology, University of Leicester, Leicester LE1 7RH, UK. ⁴Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. ⁵Geological Collections, University Museum of Natural History, Oxford OX1 3PW, UK.

the length of the lophophore (Fig. 1g). The longest, at the anterior, are about 0.8 mm long. The angle at which the tentacles are directed changes along the length of the lophophore (Fig. 1j),

The specimen bears an *in vivo* epifauna on both valves, comprising three biconvex juvenile brachiopods (OUM C.29589–29591; B1–B3 in Fig. 1a–d, h) and two encrusting organisms of uncertain affinities (OUM C.295923; E1 and E2 in Fig. 1b, c, h, k). OUM C.29589 (Fig. 1d) is probably an atrypide; it is 1.6 mm wide and transversely oval with radial plication. It preserves a short, thin pedicle, about

0.4 mm long, and marginal setae. OUM C.29590 and 29591 are too small to identify (0.5 and 0.4 mm wide, respectively). However, OUM C.29591, at least, is not conspecific with OUM C.29589 (the ?atrypide), as it is elongately oval and preserves a curved pedicle about 0.7 mm long.

Discussion. The high-level taxonomy of articulated brachiopods relies heavily on articulatory structures and on internal characters of the valves in the visceral region. In *Bethia serraticulma* these features are masked by soft tissue, and details of shell structure are also

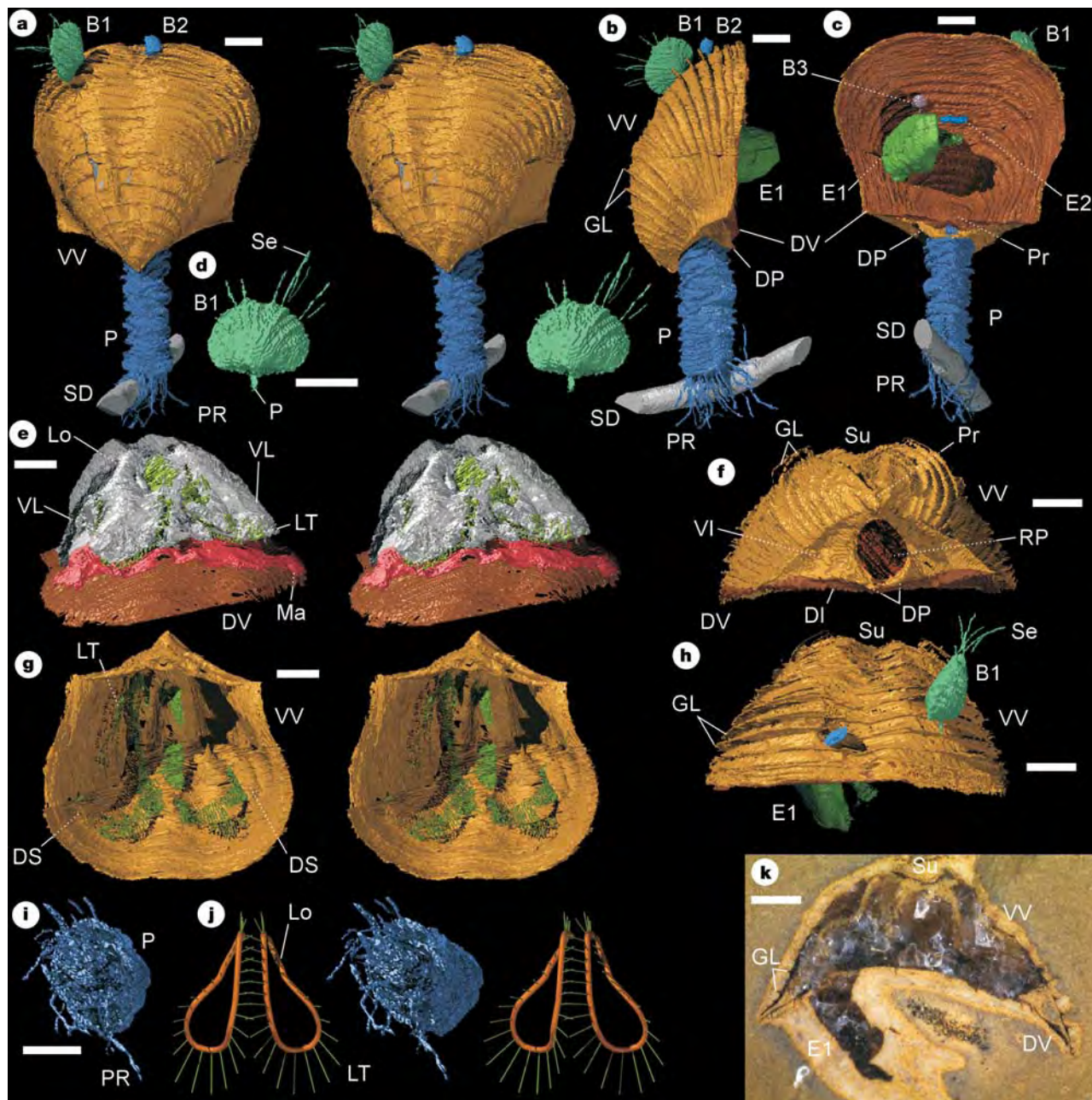


Figure 1 | *Bethia serraticulma* gen. et sp. nov. **a–c, e–i, k**, OUM C.29586; holotype of *Bethia serraticulma* (**a–c** and **e–i** are ‘virtual’ reconstructions): ventral stereo pair (**a**); right lateral view (**b**); dorsal view (**c**); subanterior internal stereo pair, ventral valve removed (**e**); posterior view, pedicle and epifauna removed (**f**); dorsal stereo pair of internal of ventral valve, all structures removed except lophophore tentacles (**g**)—note that the removed ‘internal mass’ may incorporate hard-part structures, so the simple internal valve morphology may be an artefact; anterior view (**h**); posterior stereo pair of pedicle, skeletal debris removed (**i**); photograph before serial grinding (**k**). **d**, OUM C.29589; attached ?atrypide B1, ‘virtual’ reconstruction, dorsal stereo pair. **j**, ventral stereo-pair diagram of idealized lophophore of

B. serraticulma, tentacle sparsity and length exaggerated for clarity, posterior up. Scale bars, 1 mm. Abbreviations: B1–B3, attached brachiopods OUM C.29589–29591, respectively; DI, dorsal interarea; DP, deltidial plates; DS, ?diductor muscle scars; DV, dorsal valve; E1 and E2, encrusting organisms OUM C.29592 and 29593, respectively; GL, growth lamellae; Lo, lophophore; LT, lophophore tentacle; Ma, mantle; P, pedicle; PR, pedicle rootlets; Pr, protogulum; RP, ‘removed pedicle’ (this region, obscured by the pedicle, is interpreted as an open delthyrium); Se, setae; SD, skeletal debris (to which *B. serraticulma* is attached); Su, sulcus; VI, ventral interarea; VL, visceral lobes; VV, ventral valve.

unavailable. Thus identification, even to a higher taxon, is difficult. The concavoconvex strophic shell of *B. serraticulma* is not diagnostic of any higher taxon; this morphology is most widespread in the class Strophomenata, but convergent forms are found in other clades, including the orthides, atrypides and athyrides. The ornament of the new genus is likewise of little significance at a high taxonomic level. Ventral posterior structures are informative, however, and these argue against an affinity with most strophomenates. They are most suggestive of the order Orthida, although other stem-group positions, especially in the Clitambonitidina, cannot be excluded. All the important characters of *Bethia* can be found within the Orthida, but the new genus represents a novel and somewhat aberrant character combination. An extended discussion of affinity is provided in Supplementary Information.

B. serraticulma preserves all expected non-biomineralized structures, although decay before 'freezing' of the sediment⁵ is evident in the dorsal mantle. Setae are preserved *in situ* on OUM C.29589; their absence in *B. serraticulma* is thus interpreted as genuine. The epifauna implies that most of both valves were exposed, and the elongate object to which the specimen attached is assumed to have been subhorizontal. We thus infer that the *in vivo* orientation was close to that of Fig. 1a–c, recalling the orientation inferred for immature concavoconvex strophomenides before the assumption of a quasi-infaunal (free-lying) mode of life^{12,13}. We interpret *B. serraticulma* as convergent on this strategy, and hence as an immature specimen; immaturity is also suggested by the small size and lophophore configuration.

Lophophores are known from Cambrian linguliformean and craniformean brachiopods¹⁴, but the only putative fossil lophophore previously described from an articulated brachiopod is a poorly preserved mineralized strip in the chonetid *Archeochonetes*². Although calcified lophophore supports are not uncommon in rhynchonelliformeans, the material described here uniquely documents the three-dimensional configuration of this organ directly. It is bilobed and attached to the anterior body wall; hence it represents a schizolophe, an early developmental stage from which more complex configurations are derived ontogenetically^{12,15}. Flexure away from the commissural plane is apparent (Fig. 1j), implying that this schizolophe is late-stage, likely to develop shortly into either a zygalophe-like or ptycholophe-like form.

Certain juvenile strophomenate brachiopods possess a mineralized pedicle sheath¹⁶, which can exceptionally resemble the pedicle described here (L. E. Popov, personal communication). However, *Bethia* is not interpreted as a strophomenate, and the epifauna demonstrates that pedicles do preserve in this material. We hence interpret the structure in *Bethia* as a soft-tissue pedicle. Pedicles of fossil linguliformean brachiopods are known^{14,17–19}, but these structures are not homologous with rhynchonelliformean pedicles. No fossil rhynchonellate or strophomenate pedicles have been reported previously, and a putative pedicle-base from the allied kutorginates^{20,21} is too poorly preserved for useful comparisons. Pedicle morphology varies considerably in extant (rhynchonellate) brachiopods²², but although some are as stout as that of *B. serraticulma*, these are relatively much shorter. Many are plenipectunculate^{13,23} (that is, they possess distal rootlets), but these invariably bear holdfast papillae and attach to substrates through chemical resorption^{22,24}. The rootlets described here form a physical tether to an object, indicating that holdfast papillae with resorption capabilities might have been absent. This resorption facility may be restricted to the rhynchonellate crown-group. No structures comparable to the pedicle ridges of *B. serraticulma* are known from any other pedicle; these might represent shortening structures in a contractile pedicle, or more permanent features that helped anchor the animal in the sediment. This pedicle is outside the morphological and functional range found in extant rhynchonellates. It demonstrates the dangers of extrapolating crown-group soft-tissue morphology to the stem-group.

METHODS

Morphology was reconstructed digitally after serial grinding at 20- μ m intervals^{6,11}.

Received 16 December 2004; accepted 23 May 2005.

- Williams, A., Carlson, S. J. & Brunton, C. H. C. in *Treatise on Invertebrate Paleontology* Part H, Revised, Vol. 2 (ed. Kaesler, R. L.) 1–28 (Geol. Soc. Am., Boulder, and Univ. Kansas, Lawrence, 2000).
- Racheboeuf, P. R. & Copper, P. The mesolophe, a new lophophore type for chonetacean brachiopods. *Lethaia* **23**, 341–346 (1990).
- Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Soft-bodied fossils from a Silurian volcanoclastic deposit. *Nature* **382**, 248–250 (1996).
- Sutton, M. D., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. An exceptionally preserved vermiform mollusc from the Silurian of England. *Nature* **410**, 461–463 (2001).
- Sutton, M. D., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. A three-dimensionally preserved fossil polychaete worm from the Silurian of Herefordshire, England. *Proc. R. Soc. Lond. B* **268**, 2355–2363 (2001).
- Sutton, M. D., Briggs, D. E. G., Siveter, David J., Siveter, Derek J. & Orr, P. J. The arthropod *Offacolus kingi* (Chelicerata) from the Silurian of Herefordshire, England: computer based morphological reconstructions and phylogenetic affinities. *Proc. R. Soc. Lond. B* **269**, 1195–1203 (2002).
- Siveter, David J., Sutton, M. D., Briggs, D. E. G. & Siveter, Derek J. An ostracode crustacean with soft parts from the Lower Silurian. *Science* **302**, 1749–1751 (2003).
- Briggs, D. E. G., Sutton, M. D., Siveter, David J. & Siveter, Derek J. A new phyllocarid (Crustacea: Malacostraca) from the Silurian Fossil-Lagerstätte of Herefordshire, UK. *Proc. R. Soc. Lond. B* **271**, 131–138 (2004).
- Siveter, Derek J., Sutton, M. D., Briggs, D. E. G. & Siveter, David J. A Silurian sea spider. *Nature* **431**, 978–980 (2004).
- Orr, P. J., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Three-dimensional preservation of a non-biomineralized arthropod in concretions in Silurian volcanoclastic rocks from Herefordshire, England. *J. Geol. Soc. Lond.* **157**, 173–186 (2000).
- Sutton, M. D., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Methodologies for the visualization and reconstruction of three-dimensional fossils from the Silurian Herefordshire lagerstätte. *Paleontol. Electron.* **4** (http://paleo-electronica.org/2001_1/s2/issue1_01.htm) (2001).
- Rudwick, M. J. S. *Living and Fossil Brachiopods* (Hutchinson, London, 1970).
- Bassett, M. G. in *Autoecology of Silurian Organisms* (eds Bassett, M. G. & Lawson, J. D.) 237–263 (*Spec. Pap. Paleontol.* **32**, 1984).
- Zhang, Z., Han, J., Zhang, X., Liu, J. & Shui, D. Soft-tissue preservation in the Lower Cambrian linguloid brachiopod from South China. *Acta Palaeontol. Pol.* **49**, 259–266 (2004).
- Emig, C. C. Functional disposition of the lophophore in living Brachiopoda. *Lethaia* **25**, 291–302 (1992).
- Williams, A., Brunton, C. H. C. & MacKinnon, D. I. in *Treatise on Invertebrate Paleontology* Part H, Revised, Vol. 1 (ed. Kaesler, R. L.) 321–422 (Geol. Soc. Am., Boulder, and Univ. Kansas, Lawrence, 1997).
- Walcott, C. D. A fossil *Lingula* preserving the cast of the peduncle. *Proc. US Natl Mus.* **11**, 480 (1888).
- Davidson, T. A monograph of British fossil brachiopods. *Palaeontol. Soc. Monogr.* **4**, 103–109 (1874).
- Bartels, C. & Poschmann, M. Linguloid brachiopods with preserved pedicles: Occurrence and taphonomy (Hunsrück Slate, Lower Emsian, Kaub Formation, Rhenish Massif, SW Germany). *Metalla* **9**, 123–130 (2002).
- Rowell, A. J. & Caruso, N. E. The evolutionary significance of *Nisusia sulcata*, an early articulate brachiopod. *J. Paleontol.* **59**, 1227–1242 (1985).
- Cohen, B. J., Holmer, L. E. & Lütter, C. The brachiopod fold: a neglected body plan hypothesis. *Palaeontology* **46**, 59–65 (2003).
- Bromley, R. G. & Surlyk, F. Borings produced by brachiopod pedicles, fossil and recent. *Lethaia* **6**, 349–365 (1973).
- Richardson, J. R. in *Treatise on Invertebrate Paleontology* Part H, Revised, Vol. 1 (ed. Kaesler, R. L.) 441–462 (Geol. Soc. Am., Boulder, and Univ. Kansas, Lawrence, 1997).
- Williams, A., James, M. A., Emig, C. C., Mackay, S. & Rhodes, M. C. in *Treatise on Invertebrate Paleontology* Part H, Revised, Vol. 1 (ed. Kaesler, R. L.) 7–188 (Geol. Soc. Am., Boulder, and Univ. Kansas, Lawrence, 1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Saunders for technical assistance; T. Wright, M. Bassett, L. Holmer and L. Popov for comments; and R. Fenn, T. Hall and J. Sinclair for general assistance. This work was supported by the Leverhulme Trust, NERC and English Nature.

Author Information Reprints and permissions information is available at ngp.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.D.S. (m.sutton@imperial.ac.uk).

LETTERS

Global hotspots of species richness are not congruent with endemism or threat

C. David L. Orme¹, Richard G. Davies³, Malcolm Burgess¹, Felix Eigenbrod¹, Nicola Pickup¹, Valerie A. Olson⁴, Andrea J. Webster⁵, Tzung-Su Ding⁶, Pamela C. Rasmussen⁷, Robert S. Ridgely⁸, Ali J. Stattersfield⁹, Peter M. Bennett⁴, Tim M. Blackburn⁵, Kevin J. Gaston³ & Ian P. F. Owens^{1,2}

Biodiversity hotspots have a prominent role in conservation biology^{1–9}, but it remains controversial to what extent different types of hotspot are congruent^{4,10–14}. Previous studies were unable to provide a general answer because they used a single biodiversity index, were geographically restricted, compared areas of unequal size or did not quantitatively compare hotspot types^{1–10,12–22}. Here we use a new global database on the breeding distribution of all known extant bird species to test for congruence across three types of hotspot. We demonstrate that hotspots of species richness, threat and endemism do not show the same geographical distribution. Only 2.5% of hotspot areas are common to all three aspects of diversity, with over 80% of hotspots being idiosyncratic. More generally, there is a surprisingly low overall congruence of biodiversity indices, with any one index explaining less than 24% of variation in the other indices. These results suggest that, even within a single taxonomic class, different mechanisms are responsible for the origin and maintenance of different aspects of diversity. Consequently, the different types of hotspots also vary greatly in their utility as conservation tools.

We created a global database on the geographical distribution of the breeding ranges of all known extant bird species using an equal-area grid at a resolution comparable to 1° latitude × 1° longitude. We used this database to map the geographical distribution of three different aspects of avian diversity: overall species richness (Fig. 1a); threatened species richness (Fig. 1b); and endemic species richness (Fig. 1c). Overall species richness was defined as the total number of bird species recorded as breeding in each grid cell. Threatened species richness was the number of breeding bird species in each grid cell that were listed as threatened with extinction⁸. Endemic species were the 25% of species with the smallest geographical breeding ranges^{19,23}. We used these three indices of diversity because they can be easily replicated, are well established in the literature^{4,5,8,10} and have been assumed or predicted to show congruent patterns of spatial distribution^{5–9,21,22}.

The maps of avian diversity were used to identify hotspots of species richness, threat and endemism. We initially defined hotspots as the richest 2.5% of grid cells with respect to species richness, threat or endemism, respectively. We used 2.5% as our initial criterion because several previous analyses have shown that the richest 1–5% of land area can represent a substantial proportion of terrestrial species^{1–10}. We found that, for all three measures of avian diversity, grid cells identified as being hotspots were aggregated in a relatively small number of biogeographic regions (Table 1, Supplementary Fig. 1). Hotspots of

species richness were grouped into nine distinct biogeographic regions (Fig. 2a), whereas threat hotspots were aggregated into ten regions (Fig. 2b), and endemism hotspots were aggregated in twenty biogeographic regions (Fig. 2c).

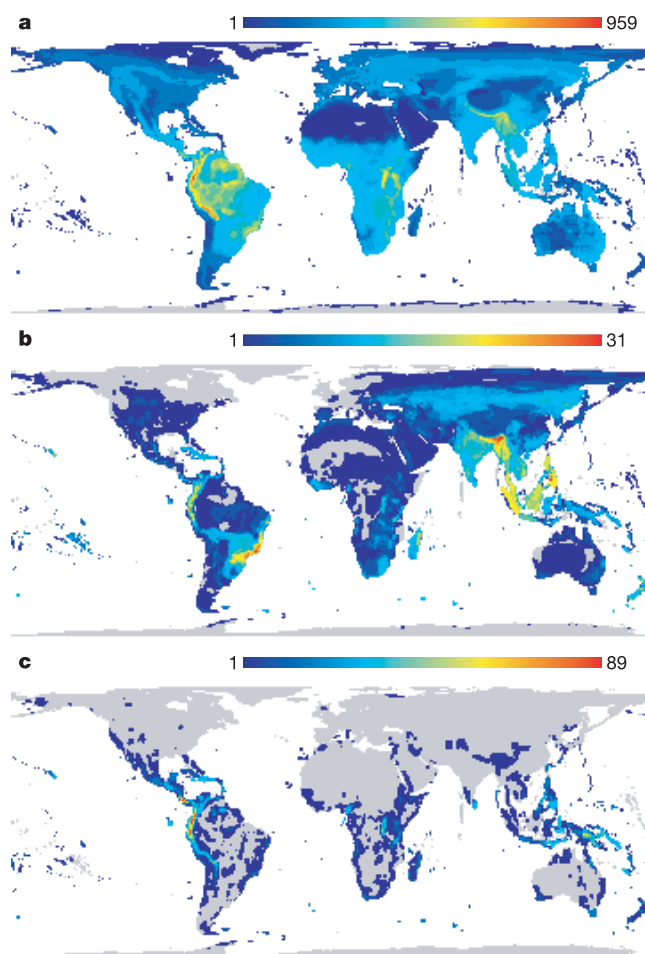


Figure 1 | Geographical distribution of three aspects of diversity. **a**, Total species richness. **b**, Threatened species richness. **c**, Endemic species richness. The bars above the maps show the corresponding colour scale, which is linear in terms of numbers of species.

¹Division of Biology and ²NERC Centre for Population Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK. ³Biodiversity and Macroecology Group, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK. ⁴Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK. ⁵School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. ⁶School of Forestry and Resource Conservation, National Taiwan University, 1, Sec 4, Roosevelt Road, Taipei 106, Taiwan. ⁷Michigan State University Museum and Department of Zoology, West Circle Drive, East Lansing, Michigan 48824-1045, USA. ⁸Academy of Natural Sciences, 1900 Benjamin Franklin Parkway, Philadelphia, Pennsylvania 19103, USA. ⁹BirdLife International, Wellbrook Court, Girton, Girton Road, Cambridge, CB3 0NA, UK.

We tested for congruence between the three types of avian hotspot by measuring the extent of spatial overlap between hotspots¹⁰. In general, there was very low spatial congruence between different types of avian hotspot. Cumulatively the three sets of hotspots occupied 1,275 grid cells, of which only 2.5% (32 grid cells) were common to all types (Fig. 3a). All of these congruent hotspot grid cells were in a single biogeographic region, the Andes (Table 1). Rather than being congruent across hotspot types, 82.4% (1,051) of hotspot grid cells were idiosyncratic to individual types, with the remaining 15.1% (192) of hotspot grid cells being shared between pairs of hotspot type (Fig. 3a). One likely reason for such low congruence was that different hotspot types were associated with different aspects of large-scale topography (Table 1). For example, of the species richness hotspot regions 89% (8 out of 9 biogeographic regions) were in mountainous areas of mainland continents, whereas only 40% (4 out of 10) and 45% (9 out of 20) of threat and endemism hotspot regions, respectively, were in continental mountains. In contrast, 60% (6 out of 10) of threat and 60% (12 out of 20) of endemism hotspot regions were on large islands and/or island archipelagos, whereas none of the species richness hotspot regions were on islands. These results agree with previous analyses that have identified highlands and islands as important regions of vertebrate diversity^{23,24}.

To verify that low congruence between avian diversity hotspots was not an artefact of our initial hotspot definition, we recalculated the extent of overlap under a variety of hotspot criteria. This showed that, even when the definition of hotspots was greatly relaxed, the

extent of overlap between the three hotspot types remained low (Fig. 3b). For example, redefining hotspots as the richest 10% of grid cells still resulted in only 4.8% (225 out of 4,664 cells) of hotspots being congruent across all three aspects of avian diversity. Low spatial overlap between different types of hotspot seems, therefore, to be a general property of global avian diversity hotspots, irrespective of the precise criterion used to define those hotspots.

Equally, the lack of congruence between different hotspot types is not an artefact of comparing only the geographical peaks of diversity, as correlations between the overall global distributions of species richness (Fig. 1a), threat (Fig. 1b) and endemism (Fig. 1c) were also surprisingly weak. Regression methods that assume that each grid cell is an independent data point are not appropriate for these data because there is strong spatial autocorrelation for all three indices of diversity (Moran's $I \geq 0.80$; $P < 0.001$ in all cases). General linear mixed models, however, can explicitly model the effects of spatial autocorrelation and were used here to explore the relationships between each pair of diversity measures. In each case, although the correlation between the two measures of diversity was statistically significant, the slope of the relationship was shallow and the proportion of variation explained was small, ranging from 7% for species endemism versus threat to approximately 22% and 24% for species richness versus threat and species richness versus endemism, respectively. These models were, therefore, consistent with our hotspot analyses and studies showing low congruence between different measures of diversity at more restricted geographical scales^{4,10,12–14,17–19}.

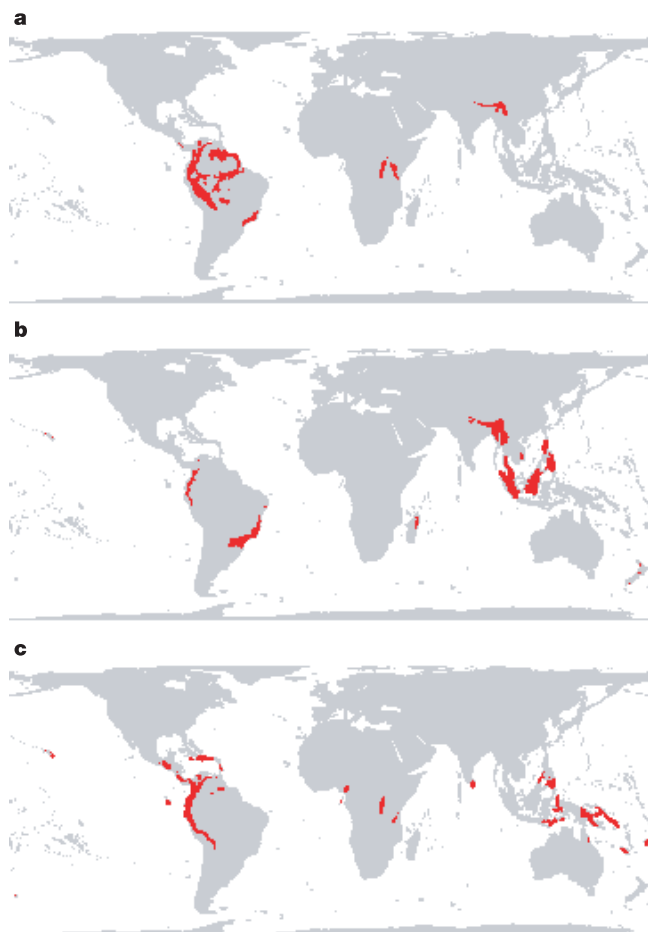


Figure 2 | Biodiversity hotspots for three aspects of diversity. **a**, Hotspots of species richness. **b**, Hotspots of threatened species. **c**, Hotspots of endemic species. For each measure of diversity, hotspots are defined as the richest 2.5% of grid cells. Hotspots are shown in red.

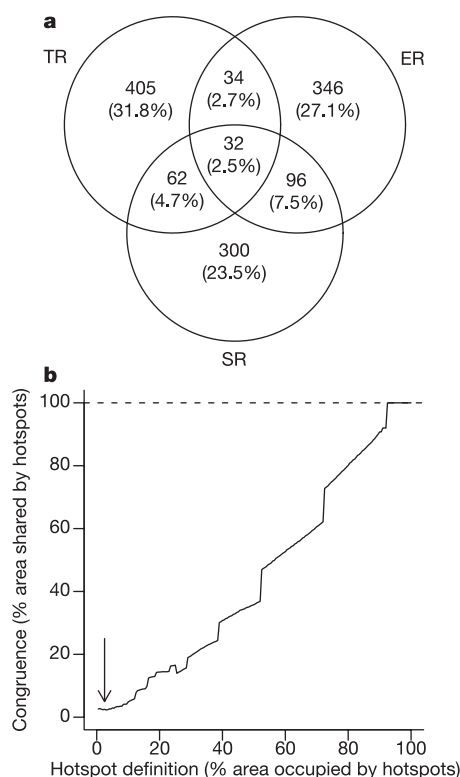


Figure 3 | Extent of congruence between hotspots. **a**, Venn diagram showing congruence across species richness hotspots (SR), threat richness hotspots (TR) and endemic richness hotspots (ER), where hotspots are the richest 2.5% of cells. Figures show number of cells and corresponding percentages. **b**, Relationship between the criterion used to define hotspots and congruence. Criteria are based on the percentage of land covered by hotspots. Congruence is the number of cells that are hotspots for all three diversity indices, as a percentage of the total hotspot area. Horizontal dashed line shows expectation under full congruence. Vertical arrow shows 2.5% hotspot criterion.

Table 1 | Avian hotspot regions with respect to species richness, threat and endemism

Hotspot regions	Type of hotspot						Topography			
	Species-richness		Threat		Endemism					
	No. spp.	Area	No. spp.	Area	No. spp.	Area	CH	CL	LI	IA
Andes	2,139	178	114	39	483	164	+			
Amazon Basin	961	105	—	—	—	—		+		
Western Great Rift Valley	936	24	—	—	60	22	+			
Eastern Great Rift Valley	902	22	—	—	19	1	+			
Himalayas	878	45	52	116	—	—	+			
Guyana highlands	877	71	—	—	32	10	+			
Atlantic coastal forests	733	27	73	95	—	—	+	+		
Mato Grosso Plateau	687	15	—	—	—	—	+			
Panama & Costa Rica highlands	621	3	—	—	101	13	+			
Philippines	—	—	47	66	68	36				+
Sumatra & Peninsula Malaysia	—	—	40	110	—	—			+	+
New Zealand	—	—	34	5	16	1			+	+
Borneo	—	—	29	82	—	—			+	
Hawaii	—	—	25	2	34	7				+
Madagascar	—	—	24	11	—	—			+	
South Vietnam highlands	—	—	19	7	—	—	+			
New Guinea & Bismarck archipelago	—	—	—	—	205	89			+	+
Caribbean	—	—	—	—	96	36				+
Lesser Sundas	—	—	—	—	72	26				+
Moluccas	—	—	—	—	68	26				+
West African forests	—	—	—	—	56	12	+			+
North Central American highlands	—	—	—	—	46	18	+			
Galapagos	—	—	—	—	31	8				+
Southern Great Rift Valley	—	—	—	—	27	7	+			
Fiji	—	—	—	—	27	9				+
Sri Lanka	—	—	—	—	23	11			+	
New Caledonia	—	—	—	—	21	9			+	+
Australian wet tropics	—	—	—	—	17	3	+			
Total	4,731	490	408	533	1,447	508				
Global	9,629	19,560	1,096	19,560	2,421	19,560				
Total as % of global	49%	2.5%	37%	2.7%	59%	2.6%				
Spatial dispersion (km)		2,079		3,023		9,469				

Figures based on hotspots defined as being the richest 2.5% of terrestrial grid cells. Hotspot regions are shown in the Supplementary Figure. No. spp. represents the cumulative number of species in the hotspot cells in a region. Area represents the number of cells identified as hotspots in a hotspot region. Large-scale topographical features contained within hotspot regions: CH, continental highlands; CL, continental lowlands; LI, large islands; IA, island archipelagos. Spatial dispersion shows the median great circle distance among hotspot cells.

Our findings of lack of congruence between different types of hotspot, and weak overall correlations among different aspects of global diversity, have important implications for understanding the ecological, evolutionary and anthropogenic mechanisms that underlie the origin and maintenance of biodiversity. It seems that, even within a single taxonomic class, different mechanisms are responsible for the geographical patterns shown by different aspects of biodiversity. This is especially intriguing in the case of species richness and endemism, which have been suggested to be closely linked in terms of overall spatial pattern^{15,16,20,25}. Our finding of a rather weak relationship between these indices agrees with the recent observation that patterns of avian species richness are determined by the distribution of widely distributed species, rather than restricted-range species¹⁹. Little is known regarding the factors determining the distribution of wide-ranging species, but a study of sub-Saharan African birds suggested important roles for energy availability²³. Endemic species richness, on the other hand, is thought to be a product of either refugia from past extinctions or of high rates of ecological and allopatric speciation^{5,23}. These observations agree with our finding that species richness hotspots are typically associated with tropical upland regions that show habitat diversity and remain forested during glacial periods, whereas endemism hotspots are more commonly on island archipelagos showing complex patterns of allopatric divergence. There have been few quantitative studies of the spatial distribution of threatened species⁸, but we predict their distribution can be determined by an interaction between the biological mechanisms promoting species diversity and the anthropogenic mechanisms eroding that diversity. We therefore expect that the lack of congruence between threat hotspots and the other two types of hotspot is due to a strong influence of human impacts on the spatial distribution of threat.

Lack of congruence among hotspot types also has implications for the use of hotspots in conservation. If congruence among hotspot types were high then it may not matter which index of diversity was used to guide conservation policy, because any such index could act as an effective surrogate for other aspects of diversity. However, our finding of very low congruence among hotspots shows that such surrogacy can not be assumed. In fact, we found that the different types of hotspot varied greatly in their ability to act as surrogates for other aspects of diversity. For instance, although the species richness hotspots (Fig. 2a) contained a large proportion of all bird species (49%, 4,731 species), they contained relatively low proportions of threatened (27%, 293 threatened species) and endemic (20%, 490 endemic species) species. In contrast, threat hotspots (Fig. 2b) contained relatively high proportions of both threatened species (37%, 408 threatened species) and overall species richness (41%, 3,932 species), but only captured a small proportion of endemic species (23%, 558 endemic species). Finally, the endemism hotspots (Fig. 2c) were successful in capturing not only a high proportion of endemic species (60%, 1,447 endemic species), but also a substantial proportion of both overall species richness (58%, 5,600 species) and threatened species richness (41%, 447 threatened species). Indeed, it is striking that the endemism hotspots actually contained a greater proportion of overall species richness than did the species richness hotspots and a greater proportion of threatened species than did the threat hotspots. These patterns need to be explored in other taxa to establish their generality, but our avian analyses indicate that endemism appears to display unusual properties, being difficult to capture using alternative indices of diversity and yet itself providing an effective way of capturing those other aspects of diversity. We suggest that these unusual properties of endemism are due to the fact that endemism hotspots are significantly more widely dispersed than

the other hotspot types (Table 1; Kruskal–Wallis test: $X^2 = 525.9$; $P < 0.001$) and therefore contain a more complementary set of species¹². This scenario provides some support for the use of endemism as a criterion for identifying hotspots^{1–3,5–7,9}, but more generally our results indicate the need to use multiple indices of diversity in identifying areas of high conservation priority^{6,7,9}.

METHODS

Mapping and hotspots. Analyses were drawn from a database of vector range maps for 9,626 extant, recognized bird species following a standard avian taxonomy²⁶. Species considered extinct were excluded from the database⁸. We mapped breeding ranges as vector maps in a geographical database using a variety of published sources (see Supplementary Methods). Vector maps were converted to a grid using a Behrmann projection at a cell resolution of 96.486 km, equivalent to 1° longitude and approximately 1° latitude at the equator (1/360th of the width of the globe under a Behrmann projection using the WGS84 datum). The global grid contained 360 by 152 cells, omitting the partial cells at latitudes higher than 87.13°. Species were scored as present in a grid cell if any of the sources indicated that the breeding range fell within the cell boundaries. Threatened species were those classified as critical, endangered or vulnerable⁸. Where necessary, we converted the taxonomy used in ref. 8 back to the standard avian taxonomy²⁶. Our definition of endemic species identified 2,421 species with ranges restricted to fewer than 30 grid cells. Terrestrial cells were defined as those containing any land from the Environmental Systems Research Institute (ESRI) digital chart of the world²⁷. Hotspot definitions were based on the percentage of terrestrial cells covered, and where quantile values fell within a richness class we used the upper number of cells for that class. The actual percentage of cells used was therefore sometimes slightly greater than target value (see Table 1 for examples at the 2.5% level). Counts of species in hotspot regions accounted for species in multiple cells.

Statistical analysis. Moran's I values were calculated using eight neighbouring cells, with P -values estimated using 1,000 randomizations: species richness = 0.95, threatened species richness = 0.91 and endemic species richness = 0.80. Generalized linear mixed effects models²⁸ used Poisson errors and fitted both predictor diversity measure and land area within each cell as fixed effects. Spatial structure was modelled using exponential covariance structures; separate range parameters were included for each of eight global biogeographic realms²⁹. To reduce computation running time to a few days for each model, models were fitted to a regular 50% subset of the data set. For the species richness versus endemism model we excluded the two smaller realms²⁹ (Antarctica and Oceania) because they prevented convergence. Estimates of the proportion of variance explained (pseudo- r^2) were computed as percentage change in total deviance of non-spatial Poisson error models. Species richness as a predictor of threat richness: slope estimate = 0.0032 (s.e. $\pm$ 0.0001; $F_{1, 9002} = 1,466.44$; $P < 0.0001$; pseudo- $r^2 = 0.218$). Species richness as a predictor of endemic species richness: slope estimate = 0.006 (s.e. $\pm$ 0.0004; $F_{1, 8699} = 190.58$; $P < 0.0001$; pseudo- $r^2 = 0.236$). Endemic species richness as a predictor of threat richness: slope estimate = 0.030 (s.e. $\pm$ 0.001; $F_{1, 9002} = 904.37$; $P < 0.0001$; pseudo- $r^2 = 0.070$). For each hotspot type, spatial dispersion was measured by first calculating for each hotspot cell the median great circle to all other hotspot cells of the same type, and then taking the median value across all hotspot cells. Median distance among endemism hotspot cells was significantly longer than corresponding distances for both species richness and threat hotspot cells (Wilcoxon tests: $P < 0.0001$ in all cases).

Received 1 April; accepted 25 May 2005.

- Myers, N. Threatened biotas: 'hotspots' in tropical forests. *Environmentalist* **8**, 187–208 (1988).
- Myers, N. The biodiversity challenge: expanded hotspots analysis. *Environmentalist* **10**, 243–256 (1990).
- Bibby, C. *et al.* *Putting Biodiversity on the Map: Priority Areas for Global Conservation* (ICPB, Cambridge, UK, 1992).
- Dobson, A. P., Rodriguez, J. P., Roberts, M. W. & Wilcove, D. S. Geographic distribution of endangered species in the United States. *Science* **275**, 550–553 (1993).
- Stattersfield, A. J., Crosby, M. J., Long, A. J. & Wege, D. C. *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation* (BirdLife International, Cambridge, UK, 1998).
- Mittermeier, R. A., Myers, N., Gil, P. R. & Mittermeier, C. G. *Hotspots: Earth's*

Biologically Richest and Most Endangered Terrestrial Ecoregions (Cemex, Conservation International and Agrupación Sierra Madre, Monterrey, Mexico, 1999).

- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
- BirdLife International, *Threatened Birds of the World* (Lynx Edicions and BirdLife International, Barcelona and Cambridge, UK, 2000).
- Mittermeier, R. A. *et al.* *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions* (Cemex, Conservation International and Agrupación Sierra Madre, Monterrey, Mexico, 2005).
- Prendergast, J. R., Quinn, R. M., Lawton, J. H., Eversham, B. C. & Gibbons, D. W. Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* **365**, 335–337 (1993).
- Pressey, R. L., Humphries, C. J., Margules, C. R., Vane-Wright, R. I. & Williams, P. H. Beyond opportunism: key principles for systematic reserve selection. *Trends Ecol. Evol.* **8**, 124–128 (1993).
- Williams, P. *et al.* A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British birds. *Conserv. Biol.* **10**, 155–174 (1996).
- Virolainen, K. M. *et al.* Hot spots, indicator taxa, complementarity and optimal networks of taiga. *Proc. R. Soc. Lond. B* **267**, 1143–1147 (2000).
- Bonn, A., Rodriguez, A. S. L. & Gaston, K. J. Threatened and endemic species: are they good indicators of patterns of biodiversity on a national scale? *Ecol. Lett.* **5**, 733–741 (2002).
- Thomas, C. D. & Mallorie, H. C. Rarity, species richness and conservation: butterflies of the Atlas Mountains in Morocco. *Biol. Conserv.* **33**, 95–117 (1985).
- Berg, A. & Tjernerberg, M. Common and rare Swedish vertebrates — distribution and habitat preferences. *Biodivers. Conserv.* **5**, 101–128 (1996).
- Howard, P. C. *et al.* Complementarity and the use of indicator groups for reserve selection in Uganda. *Nature* **394**, 472–475 (1998).
- Lawton, J. H. *et al.* Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* **391**, 72–76 (1998).
- Lennon, J. J., Koleff, P., Greenwood, J. J. D. & Gaston, K. J. Contribution of rarity and commonness to patterns of species richness. *Ecol. Lett.* **7**, 81–87 (2004).
- Jetz, W., Rahbek, C. & Colwell, R. K. The coincidence of rarity and richness and the potential signature of history in centres of endemism. *Ecol. Lett.* **7**, 1180–1191 (2004).
- Brooks, T. *et al.* Toward a blueprint for conservation in Africa. *Bioscience* **51**, 613–624 (2001).
- Baillie, J. E. M., Hilton-Taylor, C. & Stuart, S. N. *2004 IUCN Red List of Threatened Species: A Global Species Assessment* (IUCN, Gland, Switzerland and Cambridge, UK, 2004).
- Jetz, W. & Rahbek, C. Geographic range size and determinants of avian species richness. *Science* **297**, 1548–1551 (2002).
- Rahbek, C. & Graves, G. R. Multiscale assessment of patterns of avian species richness. *Proc. Natl Acad. Sci. USA* **98**, 4534–4539 (2001).
- Gaston, K. J. & Blackburn, T. M. The spatial distribution of threatened species: macro-scales and New World birds. *Proc. R. Soc. Lond. B* **263**, 235–240 (1996).
- Sibley, C. G. & Monroe, B. L. *Distribution and Taxonomy of the Birds of the World* (Yale Univ. Press, New Haven, 1990).
- Defense Mapping Agency (Mapping Agency, Fairfax, Virginia, USA, 1992).
- SAS (SAS Institute, Cary, North Carolina, 2001).
- Olson, D. M. *et al.* Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* **51**, 933–938 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Beehler, L. Birch, T. Brooks, B. Coates, J. Cromie, J. Diamond, J. Eckstrom, S. Elworthy, H. Fry, E. Harvey, A. Headon, P. Higgins, D. McNicol, D. Mehlman, C. Perrins, J. Price, R. Porter, H. Pratt, R. Prys-Jones, N. Redman, C. Robertson, B. Sheldon, I. Sherman, R. Simpson, I. Sinclair, A. Silcocks, M. Strange, M. Unwin, E. Urban, M. Weston, M. Whitby, P. Williams, D. Wynn, B. Young, J. Zook, A. and C. Black. We also thank Academic Press, BirdGuides, BirdLife International, Birds Australia, C. Helm, Conservation International, NatureServe, Oxford Univ. Press, the Ornithological Society of New Zealand, Princeton Univ. Press, the Alexander Library (Oxford), and the Natural History Museum (Tring) for access to data and libraries; and M. Balman, S. Butchart, T. Brooks, M. Cardillo, W. Jetz, A. Phillimore, H. Possingham, A. Purvis and W. Sechrest for comments or help. This work was funded by The Natural Environment Research Council.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.P.F.O. (i.owens@imperial.ac.uk).

LETTERS

Local translation of RhoA regulates growth cone collapse

Karen Y. Wu^{1*}, Ulrich Hengst^{1*}, Llewellyn J. Cox¹, Evan Z. Macosko¹, Andreas Jeromin^{2†}, Erica R. Urquhart¹ & Samie R. Jaffrey¹

Neuronal development requires highly coordinated regulation of the cytoskeleton within the developing axon. This dynamic regulation manifests itself in axonal branching, turning and pathfinding, presynaptic differentiation, and growth cone collapse and extension. Semaphorin 3A (Sema3A), a secreted guidance cue that primarily functions to repel axons from inappropriate targets, induces cytoskeletal rearrangements that result in growth cone collapse¹. These effects require intra-axonal messenger RNA translation. Here we show that transcripts for RhoA, a small guanosine triphosphatase (GTPase) that regulates the actin cytoskeleton, are localized to developing axons and growth cones, and this localization is mediated by an axonal targeting element located in the RhoA 3' untranslated region (UTR). Sema3A induces intra-axonal translation of RhoA mRNA, and this local translation of RhoA is necessary and sufficient for Sema3A-mediated growth cone collapse. These studies indicate that local RhoA translation regulates the neuronal cytoskeleton and identify a new mechanism for the regulation of RhoA signalling.

Studies using *Xenopus* retinal axons demonstrated that cytoskeletal regulation of the growth cone by Sema3A requires intra-axonal, or 'local', mRNA translation². Sema3A treatment results in increased protein synthesis in growth cones as shown by metabolic labelling experiments and by phosphorylation of elongation factor 4E-BP1 (ref. 2). These effects occur within minutes of Sema3A application². Furthermore, Sema3A-mediated growth cone collapse is blocked by ribosomal inhibitors. The mRNA translation that is required for Sema3A-mediated growth cone collapse occurs in the axon, because both Sema3A-induced collapse and inhibition of this collapse by ribosomal inhibitors is preserved in axons that are severed from their cell bodies².

To determine whether intra-axonal mRNA translation is required for Sema3A signalling in mammalian neurons, we examined Sema3A-mediated growth cone collapse in embryonic rat dorsal root ganglia (DRG) explant cultures^{3–5}. To eliminate the possibility that the effects of Sema3A were mediated through somatic translation, axons were severed from their cell bodies² (see Supplementary Fig. 1a). Treatment of severed axons with Sema3A for 60 min resulted in an increase in collapsed growth cones from $17 \pm 1.3\%$ to $75 \pm 2.8\%$ (see Supplementary Fig. 1b, c, g). This effect was blocked by pre-treatment of axons with either cycloheximide or anisomycin (see Supplementary Fig. 1), both of which are ribosomal inhibitors. Pre-treatment of these cultures with rapamycin, an inhibitor of cap-dependent translation⁶, also blocked Sema3A-mediated growth cone collapse. Together, these data indicate that the requirement for mRNA translation in mediating the cytoskeletal effects of Sema3A is a conserved feature among vertebrates.

Cytoskeletal alterations in axons are thought to involve members

of the Rho family of small GTPases, with roles for Cdc42 and Rac1 in regulating filopodia and lamellipodia formation in growth cones, and RhoA in triggering growth cone collapse and neurite retraction¹. In DRG neurons, growth cone collapse in response to Sema3A requires the RhoA effector ROCK⁷ (see Supplementary Fig. 1g), implying a role for RhoA activation in Sema3A signalling. Thus, local translation of RhoA or its effectors may contribute to Sema3A-mediated collapse. To identify axonal mRNAs that mediate the effects of Sema3A, we prepared DRG explant cultures and screened axons with a panel of probes using *in situ* hybridization. Transcripts for RhoA, as well as β -actin (an mRNA previously shown to be localized to the axon⁸), were present in axons at similar levels by fluorescent *in situ* hybridization (Fig. 1a–d; see also Supplementary Fig. 2a, b). RhoA mRNA localization was specific because transcripts for ROCK (Fig. 1e, f; see also Supplementary Fig. 2e, f), mDia (another RhoA effector; see Supplementary Fig. 2g, h), as well as other Rho family GTPases (Fig. 1g; see also Supplementary Fig. 2i–l) were not detectable in axons. We next examined RhoA mRNA localization in other types of developing neurons. *In situ* hybridization experiments with embryonic day (E)18 rat hippocampal neurons and postnatal mouse basal pontine explant neurons also showed axonal localization of RhoA transcripts (see Supplementary Fig. 2m–p). To further confirm these findings, we performed polymerase chain reaction with reverse transcription (RT–PCR) on RNA obtained from mechanically-isolated DRG axons prepared in modified Boyden chambers (see Supplementary Fig. 3a–c). RT–PCR of axonal RNA showed the presence of RhoA and β -actin transcripts, but not transcripts for ROCK1 or Rac1 (Fig. 1i). The glial-specific transcript, glial fibrillary acidic protein (GFAP), and the neuronal soma-restricted transcript γ -actin⁹, were also absent, confirming the purity of the axonal preparations (see Supplementary Fig. 3d). These data suggest that the axonal localization of RhoA mRNA is a general feature of developing axons.

In situ hybridization revealed that RhoA transcripts were localized in axons in punctate structures (Fig. 1c; see also Supplementary Fig. 2a). This localization may reflect the incorporation of RhoA mRNA in 'RNA granules' (mobile macromolecular complexes of ribosomes, fragile X-mental retardation protein (FMRP) and mRNA¹⁰) that may function to transport mRNA or as sites of translation in neurites¹¹. Indeed, staufen, a marker for RNA granules^{12,13}, partially colocalized with RhoA-transcript-containing puncta (see Supplementary Fig. 2q). We examined RhoA transcript localization in growth cones using actin immunofluorescence to define cellular borders (Fig. 1j). RhoA mRNA was found in the growth cone periphery indicating that RhoA mRNA may have a role in the regulation of the growth cone cytoskeleton.

Targeting elements in the 3'UTR of β -actin mRNA have been implicated in the localization of β -actin transcripts to axons¹⁴. We

¹Department of Pharmacology, Weill Medical College, Cornell University, New York, New York 10021, USA. ²Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA. [†]Present Address: Center for Learning and Memory, University of Texas, Austin, Texas 78712, USA.

*These authors contributed equally to this work.

did not detect these elements in the RhoA 3'UTR. To determine whether axonal localization of RhoA mRNA depends on other targeting elements in its 3'UTR, we expressed enhanced green fluorescent protein (EGFP) transcripts with various 3'UTRs in DRG explant cultures. Cultures were infected with Sindbis pseudo-virus expressing EGFP mRNAs with the minimal viral 3'UTR (conserved sequence element, 3'CSE (ref. 15), EGFP_{3'CSE}), the 3'UTR of β -actin (EGFP_{3' β -actin}) or the 3'UTR of RhoA (EGFP_{3'RhoA}) (Fig. 2a). *In situ* hybridization with an EGFP-specific probe revealed that only EGFP_{3' β -actin} or EGFP_{3'RhoA} were localized to axons and growth cones (Fig. 2b–g; see also Supplementary Fig. 4a). The absence of EGFP_{3'CSE} in axons suggests that the observed axonal localization of the EGFP_{3' β -actin} and EGFP_{3'RhoA} is not a consequence of passive diffusion or non-specific transport from the soma. *In situ* hybridization of infected cell bodies revealed that each EGFP transcript was expressed in the soma (see Supplementary Fig. 4b), indicating that the absence of EGFP_{3'CSE} in axons is not an effect of impaired transcript synthesis or stability. These studies indicate that the RhoA 3'UTR is sufficient to target heterologous transcripts to axons and growth cones.

The enrichment of RhoA mRNA in axons and growth cones suggests that its local translation might regulate the actin cytoskeleton. The findings that polyadenylated mRNA and large numbers of

polyribosomes are present in developing vertebrate axons and growth cones^{8,9,16}, as well as evidence that axons and axonal preparations have the capacity to synthesize proteins^{17–19}, suggests that axonal translation has a functional role in neuronal development. To determine whether RhoA translation is induced in response to *Sema3A*, we monitored RhoA levels after *Sema3A* treatment. Total RhoA levels were determined by summing RhoA immunofluorescence in Z-stack images (collected in 1- μ m steps) and normalizing to the volume of the growth cone, using GAP-43 immunofluorescence to define the cellular borders. DRG explants were cultured in medium containing 125 ng ml⁻¹ nerve growth factor (NGF) (rather than 75 ng ml⁻¹) to slow the rate of growth cone collapse⁷ an additional 2 h, thus allowing images of not yet collapsed growth cones to be taken after 1 h (see Supplementary Fig. 5a, b). In severed axons, *Sema3A* treatment resulted in a 2.3 ± 0.2 -fold increase in RhoA immunofluorescence (Fig. 3a–f). This increase was blocked by pre-treatment with either anisomycin or rapamycin, indicating that protein translation is required for *Sema3A*-induced increases in RhoA immunofluorescence. This increase was specific to RhoA, as total GAP-43 immunofluorescence levels were unchanged in response to *Sema3A* treatment (Fig. 3g).

To monitor the spatial distribution of protein synthesis, we used myr-dEGFP (a destabilized EGFP containing an amino-terminal myristoylation consensus sequence that limits the diffusion of the fluorescent protein from the site of translation²⁰). The myr-dEGFP

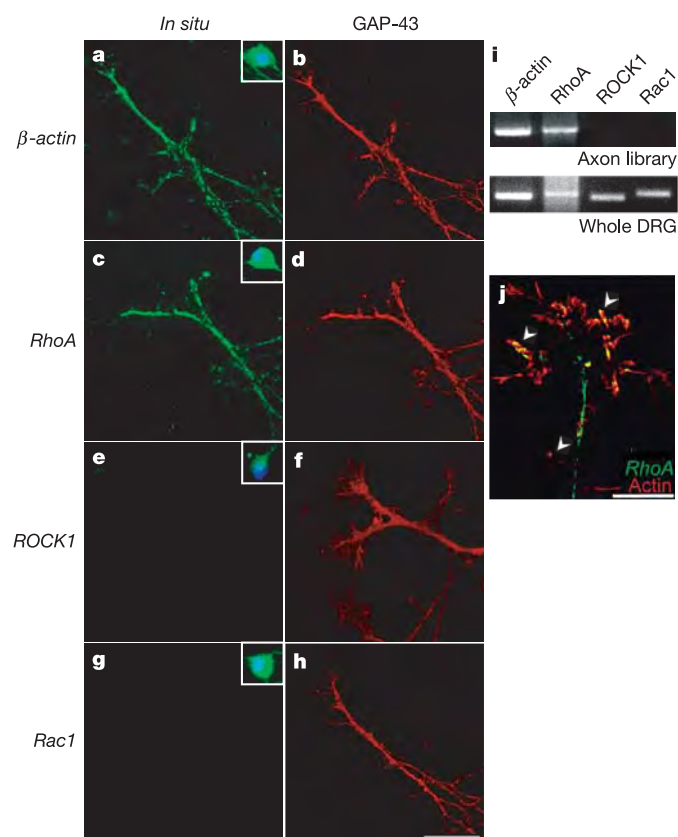


Figure 1 | RhoA mRNA is localized in axons and growth cones. **a, c, e, g,** *In situ* hybridization experiments using digoxigenin-labelled riboprobes directed against β -actin (**a**), RhoA (**c**), ROCK1 (**e**) and Rac1 (**g**) transcripts. β -actin and RhoA transcripts are localized to axons, whereas Rac1 and ROCK1 transcripts are not detectable. Inset, *In situ* hybridization of cell bodies of dissociated DRG neurons (**a, c, e, g**), DAPI staining is shown in blue. **b, d, f, h,** GAP-43 immunofluorescence of axons. Scale bar, 20 μ m. **i,** Detection of RhoA mRNA in axons by RT-PCR from purified axons. Only transcripts for RhoA and β -actin were detected in axonal preparations (upper panel), whereas all transcripts were detected when whole explant RNA was used as a template (lower panel). **j,** Growth cones were visualized by actin immunofluorescence (red). Arrowheads indicate RhoA mRNA (green) along the periphery of the growth cone. Scale bar, 10 μ m.

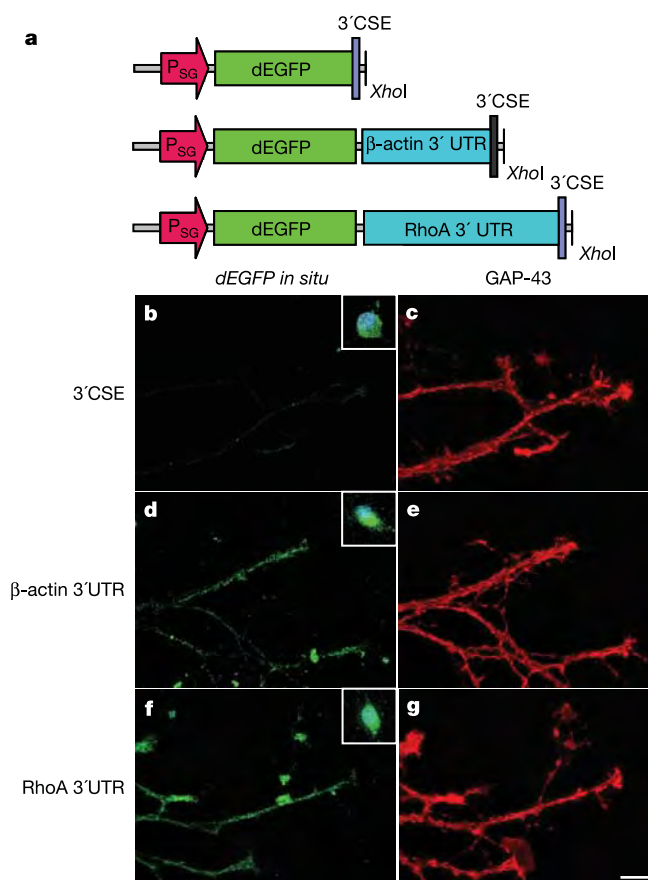


Figure 2 | The RhoA 3'UTR contains an axonal targeting element. **a,** Schematic diagram of Sindbis constructs. Transcripts encoding destabilized EGFP (dEGFP) contained the 3' CSE of Sindbis¹⁵ as the 3' UTR of the transcript, the 3' UTR of RhoA or the 3' UTR of β -actin. P_{SG} indicates the viral subgenomic promoter. **b, d, f,** *In situ* hybridization using riboprobes specific for dEGFP. dEGFP transcripts containing either the 3' UTR of β -actin or RhoA were targeted to axons. dEGFP transcripts containing only the viral 3' CSE sequence were not detected in axons. Inset, *In situ* hybridization of cell bodies. **c, e, g,** GAP-43 immunofluorescence of axons.

reporter has a half-life of one hour, thus fluorescent signals represent newly synthesized protein. Infection of DRG explants with a Sindbis viral construct comprising myr-dEGFP and the RhoA 3' UTR (Fig. 4a) resulted in the appearance of fluorescent puncta distributed throughout axons. Treatment with Sema3A for 60 min resulted in the increased intensity of pre-existing puncta as well as the appearance of new puncta in infected axons (Fig. 4b–d). New or increased fluorescence of pre-existing myr-dEGFP puncta were infrequent in the absence of Sema3A treatment (see Supplementary Fig. 6a, b). To test the role of protein translation in Sema3A-mediated increases in reporter expression, we pre-treated axons with anisomycin 30 min before the addition of Sema3A. Anisomycin pre-treatment resulted in a complete loss of Sema3A-mediated increase of myr-dEGFP throughout the axon (see Supplementary Fig. 6c, d). The absence of fluorescence in anisomycin-treated axons after 1 h of Sema3A treatment reflects the lability of myr-dEGFP and indicates that signals at this time point must derive from newly synthesized protein. Thus, the increase in myr-dEGFP levels on Sema3A treatment is due to increased translation. To assess the effects of Sema3A more comprehensively, 580 puncta were analysed from both the Sema3A-treated and the vehicle-treated axons; we then plotted the fluorescence signal at 0 min and 60 min (Fig. 4d). Analysis showed distinct populations of puncta after 60 min of Sema3A treatment, comprising both newly-formed puncta and puncta that displayed increased intensity after Sema3A treatment. Most puncta from vehicle-treated axons displayed minimal changes in fluorescence intensity over the 60 min period. The punctate localization may reflect translation at the sites of RNA granules. Furthermore, these experiments indicate that the RNA elements that direct Sema3A-induced translation are located in the 3' UTR of RhoA. The 3' UTR of RhoA contains several elements implicated in translational regulation, including several binding sites for microRNAs²¹, and a binding site for FMRP, an RNA-binding protein that regulates translation²² (see Supplementary Fig. 7).

We next examined the role of RhoA in Sema3A-mediated growth cone collapse. To assess the role of RhoA in Sema3A-mediated collapse, we examined the effects of *Clostridium botulinum* C3 exoenzyme, an ADP-ribosyltransferase that inactivates RhoA²³. Treatment of DRG explant cultures with C3 exoenzyme significantly

reduced Sema3A-induced collapse (see Supplementary Fig. 8), indicating a requirement for RhoA in the cytoskeletal remodelling effects of Sema3A.

We examined the requirement for axonal RhoA mRNA in Sema3A-mediated growth cone collapse. Treatment of cultured, dissociated DRG neurons with small interfering RNA (siRNA) directed against the 5' UTR of RhoA mRNA essentially abolished RhoA mRNA in axons (Fig. 5a–d). In these neurons, we replaced endogenous RhoA transcripts with heterologous RhoA mRNAs that exhibit selective localizations. Transcripts that contain the RhoA 3' UTR exhibit axonal localization, whereas transcripts that contain the viral 3' CSE are restricted to the soma (Fig. 2). These viral-encoded mRNAs lack the RhoA 5' UTR, rendering them resistant to the siRNA. Thus, we infected siRNA-treated neurons with Sindbis pseudovirus expressing EGFP–RhoA transcripts containing either

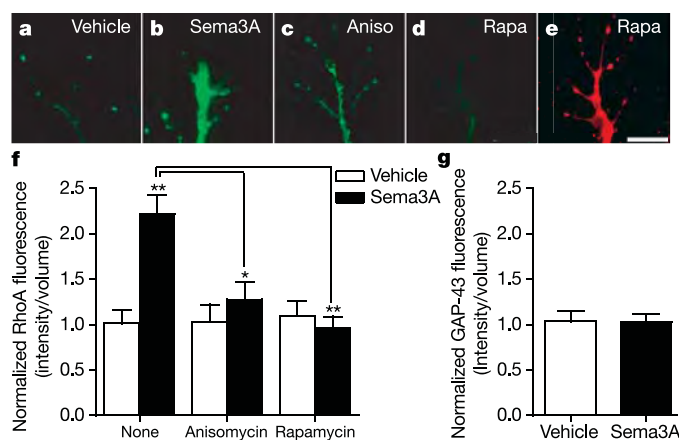


Figure 3 | Sema3A induces RhoA translation. Severed axons were pre-incubated with vehicle, 10 nM rapamycin or 40 μ M anisomycin, treated with Sema3A for 60 min, and RhoA was detected by immunofluorescence and normalized to axonal volume using GAP-43-defined cellular borders. **a**, **b**, Sema3A treatment (**b**) increases RhoA levels relative to vehicle treatment (**a**). **c**, Sema3A-induced increases in RhoA are blocked by anisomycin. **d**, Sema3A-induced increases in RhoA are blocked by rapamycin. **e**, Example of GAP-43 immunofluorescence staining of the axon terminal seen in **d**. Scale bar, 10 μ m. **f**, Summary of results from **a**–**d**. Growth cones per condition, $n = 50$; asterisk indicates $P < 0.01$ and double asterisks indicate $P < 0.001$. **g**, Normalized GAP-43 immunofluorescence is unaltered by Sema3A treatment. Growth cones per condition, $n = 50$. Error bars represent s.e.m.

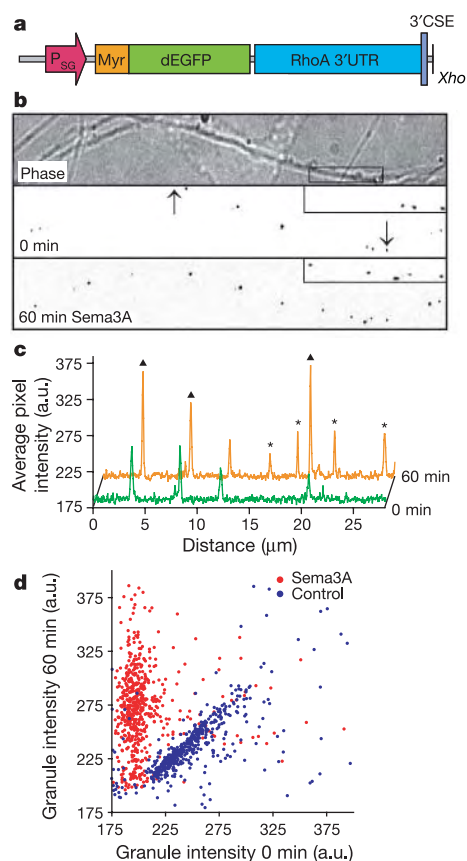


Figure 4 | Sema3A activates translation of a RhoA reporter. **a**, Schematic diagram of the Sindbis reporter construct. The reporter contains a myristoylated, destabilized EGFP (dEGFP) with the 3' UTR of RhoA. **b**, Sema3A induces translation of the reporter. Phase image (top panel) and EGFP fluorescence (middle and bottom panels) images of axons. Fluorescence images are shown with inverted contrast to facilitate visualization. Newly formed puncta and puncta with increased signal intensity are seen following Sema3A treatment (bottom panel). Inset, a region of the axon in the top panel is magnified to show changes in puncta number and intensity. Scale bar, 10 μ m. **c**, A line scan performed at 0 min (green trace) and 60 min (orange trace) in the region demarcated by arrows. Newly formed puncta (asterisks) and puncta with increased signal intensity (triangles) are seen after Sema3A treatment. **d**, Scatter plot of puncta intensities at 0 and 60 min. 580 puncta from experiments monitoring Sema3A-treated (red circles) and vehicle-treated (blue circles) axons were plotted. The majority of points from Sema3A-treated axons remained above the diagonal, indicating newly formed puncta or pre-existing puncta that increased in intensity on Sema3A treatment. Most points from vehicle-treated axons remained along the diagonal, indicating that puncta were largely unaffected by vehicle treatment. The background levels in these images averaged approximately 185.

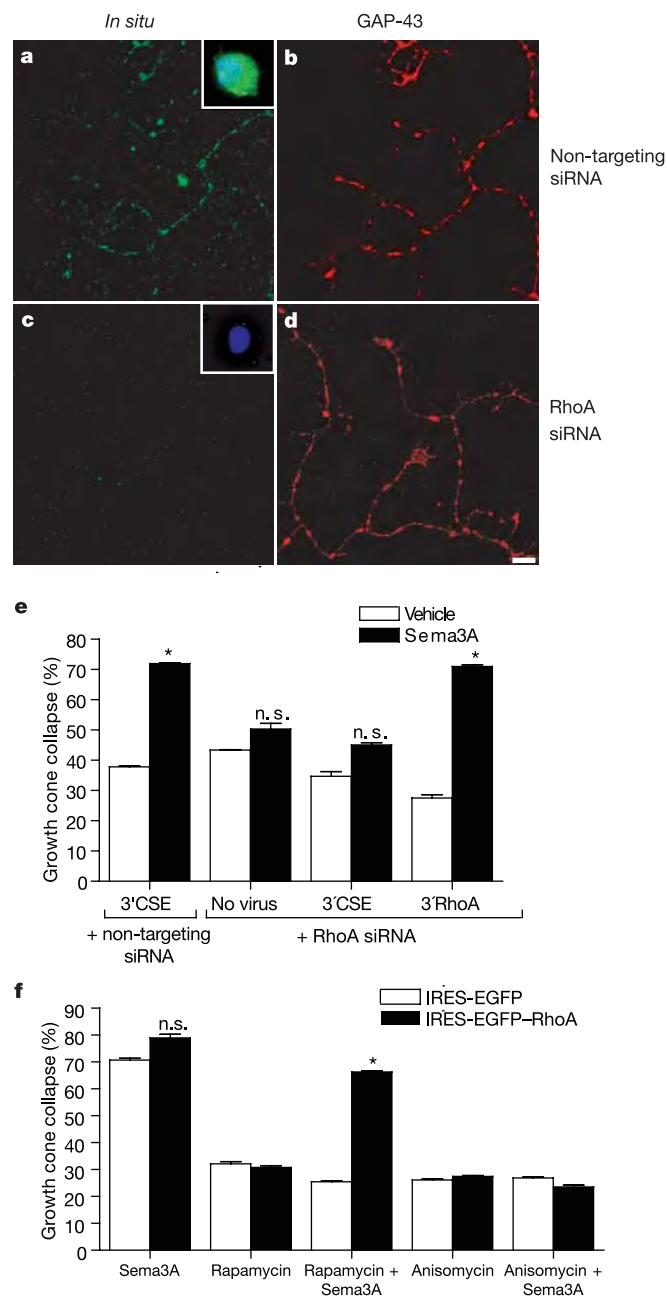


Figure 5 | Axonal RhoA translation mediates Semaphorin 3A signalling. **a–d**, siRNA-mediated knockdown of RhoA transcripts in axons. *In situ* hybridization of RhoA transcripts in DRG axons was unaffected by control siRNA (**a**) but abolished in neurons transfected RhoA 5' UTR-directed siRNA (**c**). GAP-43 immunofluorescence of axons (**b**, **d**). Inset, cell body staining. Scale bar, 10 μ m. **e**, Axonal RhoA transcripts are required for Semaphorin 3A-induced collapse. Endogenous RhoA was knocked down with RhoA 5' UTR-directed siRNA, and RhoA was restored through Sindbis pseudoviruses expressing EGFP–RhoA_{3' CSE} or EGFP–RhoA_{3' RhoA}. The domain targeted by the siRNA is absent from the viral constructs. Semaphorin 3A-mediated collapse was significantly reduced in RhoA siRNA-transfected neurons infected with pseudovirus expressing EGFP–RhoA_{3' CSE} but was restored in RhoA siRNA-transfected neurons infected with pseudovirus expressing EGFP–RhoA_{3' RhoA}. Growth cones per condition, $n = 50$; asterisk indicates $P < 0.01$. **f**, Semaphorin 3A-mediated growth cone collapse was blocked by rapamycin (10 nM) in axons expressing IRES-EGFP and restored in axons expressing IRES-EGFP–RhoA. Growth cones per condition, $n = 50$; asterisk indicates $P < 0.01$. Error bars represent s.e.m.

the viral 3' CSE (EGFP–RhoA_{3' CSE}) or the RhoA 3' UTR (EGFP–RhoA_{3' RhoA}). In DRG neurons, siRNA directed against the 5' UTR of RhoA markedly reduced Semaphorin 3A-mediated growth cone collapse (Fig. 5e). Infection of these neurons with Sindbis pseudovirus that express EGFP–RhoA_{3' CSE} failed to restore Semaphorin 3A-mediated collapse. Infection with Sindbis pseudovirus that express EGFP–RhoA_{3' RhoA} restored Semaphorin 3A-mediated growth cone collapse (Fig. 5e).

We next examined the sufficiency of RhoA translation in Semaphorin 3A-mediated growth cone collapse. Rapamycin, which blocks Semaphorin 3A-mediated growth cone collapse, blocks cap-dependent translation. However, rapamycin does not affect cap-independent translation, such as translation initiated at an internal ribosome entry site (IRES)²⁴. Thus, we generated Sindbis viral constructs that expressed RhoA in axons via an encephalomyocarditis virus IRES (see Supplementary Fig. 9). We monitored growth cone collapse in severed axons from DRG explant cultures infected with either IRES-EGFP or IRES-EGFP–RhoA (Fig. 5f). Semaphorin 3A-mediated growth cone collapse in neurons infected with IRES-EGFP (Fig. 5f) was similar to that observed in uninfected neurons (see Supplementary Fig. 1g), and was blocked by rapamycin. We next monitored growth cone collapse in DRG neurons infected with IRES-EGFP–RhoA, that permits EGFP–RhoA translation in the presence of rapamycin. We found that rapamycin treatment failed to block Semaphorin 3A-mediated growth cone collapse in neurons that express IRES-EGFP–RhoA (Fig. 5f), indicating that expression of EGFP–RhoA is sufficient to restore growth cone collapse in the presence of rapamycin. Additionally, in neurons that express IRES-EGFP–RhoA but are not treated with Semaphorin 3A, baseline levels of growth cone collapse are not significantly different from IRES-EGFP-infected neurons, indicating that EGFP–RhoA expression alone is not sufficient for collapse. Semaphorin 3A signalling probably involves both translation of RhoA and activation of the newly synthesized protein, perhaps by a Semaphorin 3A-regulated guanine nucleotide-exchange factor (GEF). The ability of EGFP–RhoA translation alone to overcome rapamycin blockade of Semaphorin 3A suggests that translation of RhoA is required for Semaphorin 3A-mediated collapse.

In summary, we have found that RhoA mRNA is enriched in developing axons, and Semaphorin 3A-regulated local translation of this transcript mediates cytoskeletal rearrangements in growth cones. Thus, regulation of mRNA translation is an effector pathway of Semaphorin 3A, probably through its receptor plexin-A²⁵. Given the prominent axonal localization of RhoA mRNA in diverse neuronal types, regulated local RhoA translation is likely to be a widespread mechanism involved in many aspects of neuronal morphogenesis. Local translation permits epistatic regulation of RhoA signalling (distinct from regulation achieved by canonical GEF pathways) by restricting activation of RhoA and its downstream effectors to sites of RhoA translation. GEFs for Rho family GTPases have been identified that mediate cytoskeletal remodelling in axons but do not display high selectivity for specific Rho family members²⁶. Activation of these GEFs may result in different outcomes that depend on whether RhoA translation pathways have been activated.

METHODS

Embryonic day (E)15–16 dissociated DRG neurons²⁷ and explants were plated on glass pre-coated with 33 μ g ml⁻¹ poly-D-lysine and 1 μ g ml⁻¹ fibronectin. DRGs were cultured in B27/F-12/MEM (Invitrogen) supplemented with 75 ng ml⁻¹ NGF, 40 mM glucose, 10 μ M 1-(β -D-arabinofuranosyl)cytosine and 20 μ M 5-fluorodeoxyuridine. RhoA 5' UTR-directed siRNA (target sequence: 5'-AAU-GAGCCUUGCAUCUAAGAA-3') was transfected using GeneSilencer (Genlantis) according to the manufacturer's protocol. siRNA transfection was performed at 2 days *in vitro* (DIV), 1 d after viral infection. For RhoA rescue, EGFP–RhoA viruses were infected at titers giving equivalent EGFP cell-body fluorescence.

Cycloheximide, anisomycin, rapamycin or *C. botulinum* C3 exoenzyme was bath-applied 30 min before the application of 450 ng ml⁻¹ Semaphorin 3A (R&D) or vehicle (0.1% BSA in PBS) for 60 min in collapse assays. This concentration of Semaphorin 3A was empirically determined to cause the maximal degree of collapse within 60 min. Only axons without varicosities or blebbing and whose growth cones remained healthy for the duration of the experiment were evaluated.

Phase-contrast images of the same growth cones were taken at 0 and 60 min. A growth cone was considered collapsed if it had less than two filopodia, each shorter than $10\ \mu\text{m}^2$. Values are presented as per cent collapsed growth cones $\pm$ s.e.m. and *P* values were determined using the student *t*-test from experiments repeated a minimum of three times.

For measurement of rapamycin-resistant collapse, DIV2 DRGs were infected with IRES viruses and collapse assays were performed 48 h after infection. Boyden chambers²⁸ and RT-PCR²⁹ were prepared with modifications as described in Supplementary Fig. 3.

In situ hybridization. Antisense riboprobes were transcribed *in vitro* from sense oligonucleotides (see Supplementary Table) using MEGAscript (Ambion) with digoxigenin-conjugated UTP.

For fluorescent *in situ* hybridizations (FISH), DRGs were fixed overnight at 4 °C in 4% paraformaldehyde (PFA) in cytoskeleton buffer (CSB: 10 mM MES at pH 6.1, 138 mM KCl, 3 mM MgCl₂, 2 mM EGTA and 0.4 M sucrose). Unless indicated, washes were performed in TBST (20 mM Tris at pH 8.0, 150 mM NaCl and 0.1% Triton X-100) for 3 × 5 min. DRGs were permeabilized (0.5% Triton X-100 in TBS) for 10 min, then post-fixed (4% PFA in TBS) for 5 min, followed by acetylation (0.25% acetic anhydride in 0.1 M HEPES) for 10 min before equilibration with 4 × sodium chloride-sodium citrate (SSC)/50% formamide for 20 min. For hybridization, coverslips were incubated with 15 ng riboprobes in 15 μl hybridization buffer (10% dextran sulphate, 4 × SSC, 1 × Denhardt's Solution, 40% formamide, 20 mM ribonucleoside vanadyl complex, 10 mM dithiothreitol, 1 mg ml⁻¹ yeast tRNA and 1 mg ml⁻¹ salmon sperm DNA) at 37 °C overnight. The coverslips were washed with 40% formamide/1 × SSC at 37 °C for 20 min, three times each with 1 × SSC and 0.1 × SSC. DRGs were blocked (100 mM Tris-HCl at pH 8.0, 150 mM NaCl, 8% formamide, 5% BSA, 2.5% normal horse serum and 2.5% normal goat serum) for 30 min. Anti-digoxin antibody (1:500, DI-22, Sigma) was pre-cleared with rat embryo powder for 2 h at 25 °C. Antibodies were as follows: GAP-43, GFAP, staufen (Chemicon); tau (Sigma); actin, RhoA (Santa Cruz Biotechnology). Secondary antibodies used were Alexa Fluor 488 and 546 (Molecular Probes).

Image analysis and quantification of immunofluorescence. DRGs were fixed with 4% PFA in CSB overnight at 4 °C, permeabilized with 0.5% Triton X-100 in TBS, and blocked in 2% BSA in 20 mM Tris-HCl at pH 8.0 for 30 min. DRGs were labelled with anti-RhoA (26C4, 1:1,000). Similar results were obtained with a RhoA antibody against a separate epitope (sc-179, 1:1,000).

For image acquisition of RhoA mRNA (*in situ* hybridization), RhoA immunofluorescence or GAP-43 immunofluorescence, exposure times were kept constant and below grey scale saturation. Three-dimensional deconvolution was done with AutoDeblur (AutoQuant).

For RhoA immunofluorescence normalization, the signal in the GAP-43 immunofluorescence stack was thresholded and used to create a volume mask around fluorescent objects. This mask was then applied onto the RhoA Z series and the total pixel intensity within this volume was measured. The summed immunofluorescence intensity was normalized to the volume of the growth cone. Quantification of *in situ* hybridization was performed similarly, using the GAP-43 labelling to demarcate cellular boundaries.

For quantification of myr-DEGFP puncta a stack of images was three-dimensionally deconvoluted. The stack of images was collapsed to a single image using the maximal projection. The fluorescence signal was thresholded, a top-hat filter was applied and regions around the fluorescent puncta were automatically traced. To exclude debris, a size exclusion limit was defined. The regions were transferred to the corresponding picture (0 or 60 min, respectively) and the average pixel intensities in the regions were plotted. Image analyses were performed with MetaMorph (Universal Imaging).

Generation and infection of Sindbis virus. We used a Sindbis vector, pSinRep5, containing a point mutation in nsP2 (P726S) that reduces neuronal cytotoxicity³⁰, and the helper plasmid DH-BB (S. Schlesinger, Washington University). The 3' nontranslated region (3' NTR) of pSinRep5 was replaced with 3' UTRs of interest, preserving the 29 nucleotide CSE required for replication and polyadenylation¹⁵. Pseudoviruses were prepared according to the manufacturer's instructions (Invitrogen), purified using a sucrose gradient, concentrated and titered using BHK-21 cells.

Received 11 April; accepted 6 June 2005.

- Gallo, G. & Letourneau, P. C. Regulation of growth cone actin filaments by guidance cues. *J. Neurobiol.* **58**, 92–102 (2004).
- Campbell, D. S. & Holt, C. E. Chemotropic responses of retinal growth cones mediated by rapid local protein synthesis and degradation. *Neuron* **32**, 1013–1026 (2001).
- Luo, Y., Raible, D. & Raper, J. A. Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones. *Cell* **75**, 217–227 (1993).

- Kolodkin, A. L. *et al.* Neuropilin is a semaphorin III receptor. *Cell* **90**, 753–762 (1997).
- He, Z. & Tessier-Lavigne, M. Neuropilin is a receptor for the axonal chemorepellent Semaphorin III. *Cell* **90**, 739–751 (1997).
- Beretta, L., Gingras, A. C., Svitkin, Y. V., Hall, M. N. & Sonenberg, N. Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation. *EMBO J.* **15**, 658–664 (1996).
- Dontchev, V. D. & Letourneau, P. C. Nerve growth factor and semaphorin 3A signalling pathways interact in regulating sensory neuronal growth cone motility. *J. Neurosci.* **22**, 6659–6669 (2002).
- Olink-Coux, M. & Hollenbeck, P. J. Localization and active transport of mRNA in axons of sympathetic neurons in culture. *J. Neurosci.* **16**, 1346–1358 (1996).
- Bassell, G. J. *et al.* Sorting of β-actin mRNA and protein to neurites and growth cones in culture. *J. Neurosci.* **18**, 251–265 (1998).
- Kanai, Y., Dohmae, N. & Hirokawa, N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron* **43**, 513–525 (2004).
- Krichevsky, A. M. & Kosik, K. S. Neuronal RNA granules: a link between RNA localization and stimulation-dependent translation. *Neuron* **32**, 683–696 (2001).
- Tang, S. J., Meulemans, D., Vazquez, L., Colaco, N. & Schuman, E. A role for a rat homolog of staufen in the transport of RNA to neuronal dendrites. *Neuron* **32**, 463–475 (2001).
- Kohrmann, M. *et al.* Microtubule-dependent recruitment of Staufen-green fluorescent protein into large RNA-containing granules and subsequent dendritic transport in living hippocampal neurons. *Mol. Biol. Cell* **10**, 2945–2953 (1999).
- Kislauskis, E. H., Zhu, X. & Singer, R. H. Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. *J. Cell Biol.* **127**, 441–451 (1994).
- Raju, R., Hajjou, M., Hill, K. R., Botta, V. & Botta, S. *In vivo* addition of poly(A) tail and AU-rich sequences to the 3' terminus of the Sindbis virus RNA genome: a novel 3'-end repair pathway. *J. Virol.* **73**, 2410–2419 (1999).
- Tennyson, V. M. The fine structure of the axon and growth cone of the dorsal root neuroblast of the rabbit embryo. *J. Cell Biol.* **44**, 62–79 (1970).
- Brittis, P. A., Lu, Q. & Flanagan, J. G. Axonal protein synthesis provides a mechanism for localized regulation at an intermediate target. *Cell* **110**, 223–235 (2002).
- Eng, H., Lund, K. & Campenot, R. B. Synthesis of β-tubulin, actin, and other proteins in axons of sympathetic neurons in compartmented cultures. *J. Neurosci.* **19**, 1–9 (1999).
- Steward, O. & Schuman, E. M. Compartmentalized synthesis and degradation of proteins in neurons. *Neuron* **40**, 347–359 (2003).
- Aakalu, G., Smith, W. B., Nguyen, N., Jiang, C. & Schuman, E. M. Dynamic visualization of local protein synthesis in hippocampal neurons. *Neuron* **30**, 489–502 (2001).
- John, B. *et al.* Human microRNA targets. *PLoS Biology* **2**, e363 (2004).
- Chen, L., Yun, S. W., Seto, J., Liu, W. & Toth, M. The fragile X mental retardation protein binds and regulates a novel class of mRNAs containing U rich target sequences. *Neuroscience* **120**, 1005–1017 (2003).
- Aktories, K. Rho proteins: targets for bacterial toxins. *Trends Microbiol.* **5**, 282–288 (1997).
- Svitkin, Y. V., Hahn, H., Gingras, A. C., Palmenberg, A. C. & Sonenberg, N. Rapamycin and wortmannin enhance replication of a defective encephalomyocarditis virus. *J. Virol.* **72**, 5811–5819 (1998).
- Fiore, R. & Puschel, A. W. The function of semaphorins during nervous system development. *Front. Biosci.* **8**, s484–s499 (2003).
- Shamah, S. M. *et al.* EphA receptors regulate growth cone dynamics through the novel guanine nucleotide exchange factor ephexin. *Cell* **105**, 233–244 (2001).
- Svenningsson, A. F., Shan, W. S., Colman, D. R. & Pedraza, L. Rapid method for culturing embryonic neuron-glia cell cocultures. *J. Neurosci. Res.* **72**, 565–573 (2003).
- Zheng, J. Q. *et al.* A functional role for intra-axonal protein synthesis during axonal regeneration from adult sensory neurons. *J. Neurosci.* **21**, 9291–9303 (2001).
- Tietjen, I. *et al.* Single-cell transcriptional analysis of neuronal progenitors. *Neuron* **38**, 161–175 (2003).
- Jeromin, A., Yuan, L. L., Frick, A., Pfaffinger, P. & Johnston, D. A modified Sindbis vector for prolonged gene expression in neurons. *J. Neurophysiol.* **90**, 2741–2745 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Schlesinger for Sindbis plasmids and N. O'Connor for advice on 3D deconvolution and axonal volume calculation. This work is supported by the National Institute of Mental Health (S.R.J.), the National Alliance for Autism Research (S.R.J.), the Charles A. Dana foundation and the Medical Scientist Training Program (E.Z.M.). A.J. is supported by D. Johnston (Baylor Medical College).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.R.J. (srj2003@med.cornell.edu).

Dystrophic heart failure blocked by membrane sealant poloxamer

Soichiro Yasuda^{1*}, DeWayne Townsend^{1*}, Daniel E. Michele^{1,2}, Elizabeth G. Favre¹, Sharlene M. Day² & Joseph M. Metzger^{1,2}

Dystrophin deficiency causes Duchenne muscular dystrophy (DMD) in humans, an inherited and progressive disease of striated muscle deterioration that frequently involves pronounced cardiomyopathy¹. Heart failure is the second leading cause of fatalities in DMD^{1,2}. Progress towards defining the molecular basis of disease in DMD has mostly come from studies on skeletal muscle, with comparatively little attention directed to cardiac muscle. The pathophysiological mechanisms involved in cardiac myocytes may differ significantly from skeletal myofibres; this is underscored by the presence of significant cardiac disease in patients with truncated or reduced levels of dystrophin but without skeletal muscle disease³. Here we show that intact, isolated dystrophin-deficient cardiac myocytes have reduced compliance and increased susceptibility to stretch-mediated calcium overload, leading to cell contracture and death, and that application of the membrane sealant poloxamer 188 corrects these defects *in vitro*. *In vivo* administration of poloxamer 188 to dystrophic mice instantly improved ventricular geometry and blocked the development of acute cardiac failure during a dobutamine-mediated stress protocol. Once issues relating to optimal dosing and long-term effects of poloxamer 188 in humans have been resolved, chemical-based membrane sealants could represent a new therapeutic approach for preventing or reversing the progression of cardiomyopathy and heart failure in muscular dystrophy.

It is unknown whether dystrophin deficiency directly causes altered force transmission and/or membrane fragility in cardiac muscle at the single myocyte level. Here we used a unique micro-carbon fibre technique (Supplementary Fig. 1 and Video 1) that enables the simultaneous assessment of force and intracellular calcium concentrations of isolated, membrane-intact myocytes under physiologically relevant mechanical loading. Single, membrane-intact adult cardiac myocytes from dystrophin-deficient mice (*mdx*, also known as *Dmd*^{*mdx*}) and control mice were passively stretched (1-s duration) over a physiologically relevant range of sarcomere lengths (1.75–2.20 μm ; ref. 4), and passive tension and intracellular calcium concentrations ($[\text{Ca}^{2+}]_i$) were recorded. Figure 1a shows photomicrographs of control and *mdx* myocytes at different stages of the stretch protocol. In Fig. 1b, representative simultaneous recordings of length, tension and Fura-2 ratio of single cardiac myocytes from a control and *mdx* mouse are shown. The traces on the left are the active twitch tension and calcium transient during an electrically stimulated isometric contraction at a resting sarcomere length of 1.75–1.80 μm . Remaining traces are tension and calcium recordings during passive stretching of the myocyte from a resting sarcomere length of 1.75 μm (0% stretch) to a physiologically relevant diastolic sarcomere length of 2.00–2.20 μm (20% stretch). Peak isometric twitch tension is not different between control and

mdx myocytes (3.6 ± 0.7 and $4.4 \pm 0.8 \text{ mN mm}^{-2}$, respectively; mean $\pm$ s.e.m; $P = 0.47$), indicating that excitation–contraction coupling and force generation/transmission are normal in *mdx* myocytes. However, dystrophin deficiency did have significant effects on passive physiological stretches in myocyte sarcomere length (Fig. 1b). With passive excursions from resting sarcomere lengths of 1.80 μm to longer sarcomere lengths, *mdx* myocytes developed significantly increased tension compared with control myocytes (Fig. 2a). This is in contrast to previous studies on skeletal myotubes, where dystrophic muscles were found to be more compliant than control muscles⁵. Stretches to sarcomere lengths greater than 2.10 μm resulted in *mdx* myocytes becoming unstable, with increased $[\text{Ca}^{2+}]_i$, fibrillations, eventual calcium overload and subsequent contracture and myocyte death (Fig. 1 and Supplementary Video 2). Reducing extracellular $[\text{Ca}^{2+}]$ from 1.8 to 0.2 mM shifted the tension–extension curves rightward, permitting stretches of up to 2.30 μm in both *mdx* and control myocytes (data not shown). This highlights a role for extracellular calcium in mediating the reduced compliance of *mdx* myocytes.

These results establish a primary defect in cell compliance in single, isolated *mdx* myocytes, which show increased susceptibility to stretch-mediated membrane instability and calcium-dependent hyper-contracture. We next tested whether a chemical-based membrane repair approach would have efficacy in isolated *mdx* myocytes. The nonionic triblock co-polymer poloxamer 188, poly(ethylene oxide)₈₀-poly(propylene oxide)₂₇-poly(ethylene oxide)₈₀ (molecular mass $\sim 8.4 \text{ kDa}$), which has previously been shown to insert into artificial lipid monolayers and repair damaged biological membranes^{6,7}, was assessed for its ability to stabilize *mdx* myocyte membranes during physiological loading conditions *in vitro* (Fig. 2a). Within the physiologically relevant sarcomere length range of 1.80–2.20 μm , 150 μM P188 (ref. 8) fully restored *mdx* myocyte compliance and $[\text{Ca}^{2+}]_i$ to control levels (Fig. 2). This dose of P188 was shown to be effective in protecting skeletal muscle from electrocution damage⁸. P188 also significantly improved twitch tension performance after a lengthening contraction (Supplementary Fig. 2). P188 had no effect on control C57BL/10 myocyte compliance within this sarcomere length range ($P = 0.49$ at 2.0 μm and $P = 0.58$ at 2.1 μm ; Fig. 2a).

The increase in $[\text{Ca}^{2+}]_i$ resulting from passive stretch is significantly greater in *mdx* myocytes, and this is corrected by P188 (Fig. 2b). To further address the mechanism of calcium entry, myocytes were treated with the L-type Ca^{2+} channel blocker nifedipine. In contrast to P188, heightened $[\text{Ca}^{2+}]_i$ upon stretch was not blocked by nifedipine (Fig. 2b), suggesting that elevated $[\text{Ca}^{2+}]_i$ may arise from channels that are not sensitive to dihydropyridine (DHP). Calcium leak channels are another possible pathway of calcium entry, but are not activated by stretch⁹. The most likely mechanism of

¹Departments of Molecular and Integrative Physiology and ²Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109-0622, USA.

*These authors contributed equally to this work.

calcium entry is through small tears in the membrane that are prevented by P188 in *mdx* myocytes. In additional experiments, a lower concentration of P188 (60 μM), previously shown to enhance membrane resealing in toxin-poisoned fibroblasts¹⁰, significantly improved *mdx* myocyte stability during passive stretch (Fig. 2c).

To test directly whether myocyte stretching causes membrane damage in *mdx* myocytes, the styryl dye FM 1-43 was used. This compound is impermeable to intact cell membranes, and specifically fluoresces when incorporated into the lipid environment of

cellular membranes. FM 1-43 has been used to demonstrate membrane disruption/repair in a range of cell systems¹⁰, including dysferlin-deficient skeletal muscle fibres¹¹. We subjected single cardiac myocytes to a single stretch of 20% in the presence of 2.5 μM FM 1-43 (Fig. 2d). Control C57BL/10 myocytes were mechanically stable after stretch, with no change in FM 1-43 fluorescence. In contrast, *mdx* myocytes became unstable after stretch, showing fibrillations (Supplementary Video 3) but not cell contraction (Fig. 1a, label *h*), and showing a concurrent steady increase in FM 1-43 fluorescence. Stretch-mediated *mdx* myocyte instability/fibrillation and elevated FM 1-43 fluorescence were completely blocked by P188. We hypothesize that a $\sim 20\%$ stretch causes instability of the *mdx* myocyte, allowing transmembrane influx of Ca^{2+} that triggers cycles of myocyte fibrillations, which in turn create more membrane instability. We interpret the steady increase in FM 1-43 fluorescence after the triggering stretch as the accumulation of dye incorporated into the plasma membrane owing to continuous sarcolemmal disruptions of the fibrillating myocyte. The time course of FM 1-43 fluorescence in *mdx* myocytes after stretch differs somewhat from the biphasic increase in fluorescence observed in skeletal muscles after a single laser-generated pore in the membrane¹¹. The type of inciting damage (laser focal point damage in skeletal muscle versus stretch of entire cardiac myocyte in this study) and differences in calcium handling between cardiac and skeletal muscles (calcium-induced calcium release being significantly more prominent in cardiac than skeletal muscle) probably account for the differing kinetics of FM 1-43 observed here. We also found that stretching *mdx* myocytes by more than 20% could cause fibrillations and elevated FM 1-43 fluorescence even in the presence of P188, indicating that the fluorescence properties of FM 1-43 are not altered by P188. From these results we conclude that P188 enhances sealing of the *mdx* plasma membrane sufficiently to prevent the initial stretch-induced injury from causing cycles of membrane de-stabilization, Ca^{2+} entry and fibrillation.

We next determined whether the cellular effects of P188 in preventing stretch-induced membrane damage in *mdx* cardiac myocytes would translate to preventing cardiac dysfunction in *mdx* mice *in vivo*. Baseline left ventricular haemodynamic performance was depressed in *mdx* mice, including reduced left ventricular end-diastolic volume (LVEDV) (Fig. 3 and Supplementary Table 1). Pre-treatment with P188 by intravenous infusion increased LVEDV to levels seen in control hearts. We hypothesize that the lower LVEDV in *mdx* hearts is an organ-level manifestation of the membrane defect observed in single isolated myocytes, and that the acute rescue of LVEDV by P188 in the *mdx* heart is a direct effect of restoring normal compliance in single *mdx* cardiac myocytes (Fig. 2). The haemodynamic parameters (Supplementary Table 1) indicate that the primary effect of acute P188 administration is to restore end-diastolic volumes without significantly changing end-diastolic pressure or other parameters, consistent with a role for P188 in improving the compliance of the *mdx* cardiac myocytes *in vitro* (Fig. 2).

Acute cardiomyopathy and heart failure can be induced by cardiovascular stressors in *mdx* mice¹². We therefore tested whether an acute dobutamine stress challenge *in vivo* could cause acute cardiac failure, and whether this phenotype could be blocked by P188. Untreated *mdx* mice had a very attenuated response to dobutamine infusion (data not shown) and a significant incidence of acute cardiac failure (Fig. 3c) during the 30-min stress-test regime, whereas control mice responded robustly to dobutamine infusion and did not develop acute heart failure. Pre-treatment of *mdx* mice by intravenous infusion of P188 immediately improved the haemodynamic response to dobutamine infusion (data not shown) and conferred protection from dobutamine-induced acute heart failure *in vivo* ($P = 0.005$, Fig. 3c).

Collectively, these findings demonstrate that in dystrophin-

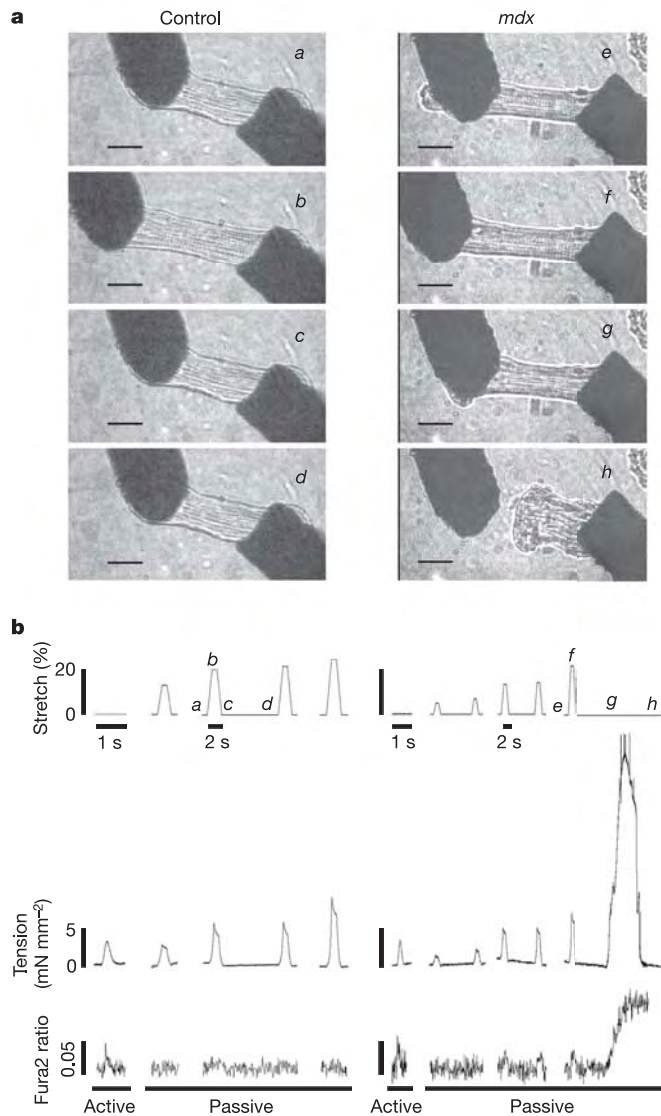


Figure 1 | Active and passive tension and $[\text{Ca}^{2+}]_i$ in single cardiac myocytes from control and *mdx* mice. Representative recordings from control (left panels) and *mdx* myocytes (right panels). **a**, Photomicrographs of single myocytes before (*a*, *e*), during (*b*, *f*), immediately after (*c*, *g*) and about 40 s after (*d*, *h*) a single passive stretch. Scale bar, 20 μm . **b**, Top traces show changes in myocyte length, starting at resting sarcomere lengths of 1.75–1.80 μm (0% stretch; isometric twitch) and extending to 2.1 μm (20% stretch and beyond). Traces marked *a*–*h* correspond to the sequence of cell stretch (*a*–*h*) shown in **a**. Middle traces show tension recordings in response to stretch. Bottom traces are $[\text{Ca}^{2+}]_i$ (Fura2 ratios) during stretches. Leftmost traces are active isometric twitches. Passive recordings are in the absence of electrical stimulus. In *mdx* mice, myocytes became unstable after stretch (*f*) and return to rest length (*g*), with massive increase in $[\text{Ca}^{2+}]_i$, hyper-contraction and death (*h*).

deficient hearts, abnormal stretch-induced increases in $[Ca^{2+}]_i$ result in decreased compliance at the cellular level and lower diastolic volume *in vivo*. The ability of the membrane sealant P188 to correct these abnormalities suggests that the calcium influx results from a loss of membrane integrity. Endogenous mechanisms of membrane repair have recently been investigated in dysferlin-deficient dystrophic skeletal muscles¹¹, and some of these repair mechanisms are elevated as part of the compensatory response in skeletal muscles of *mdx* mice¹³. It is reasonable to assume that similar mechanisms are present within the cardiomyocytes, but that these are not sufficient to redress membrane instability, particularly under heightened mechanical stress.

Current therapeutic strategies for Duchenne muscular dystrophy are focused on the expression of dystrophin (through exon skipping or viral transduction of truncated dystrophin) or other genes (for

example, utrophin or dysferlin) that limit the consequences of dystrophin deficiency^{14–17}. The present results suggest a comparatively simple, chemical-based alternative involving membrane stabilization and/or repair by intravenous administration of P188. As shown here in this animal model of DMD, P188 can have immediate beneficial haemodynamic effects under basal conditions, and may prevent myocardial damage under acute stress conditions, for example during surgery in DMD patients. Currently, P188 is in phase III clinical trials for the treatment of vaso-occlusive crises in sickle-cell anaemia patients, having recently demonstrated the safety and non-toxicity of P188 for short-term (24-h) use in humans¹⁸. Unlike the episodic course of sickle-cell anaemia, DMD is a progressive disease, and effective P188 therapy would probably require chronic intravascular administration. Long-term effects of this sealant are unclear at present and would have to be validated in large-animal models of muscular dystrophy. In one safety study designed for 72 h the highest dose tested was terminated early because of adverse effects (primarily muscle pain) in healthy volunteers, indicating that high doses may be too toxic for long-term use¹⁹. Human studies will be required to fully address the optimal dose, route and duration (acute and chronic) for delivery of chemical membrane sealants. If issues of dosing and long-term safety can be addressed, our results indicate that membrane-sealing poloxamers could represent a new class of therapeutic agents for preventing or limiting progressive damage to the hearts of DMD patients, and might also have benefits in other cardiomyopathies associated with defects in the dystrophin glycoprotein complex²⁰.

METHODS

Animals. Control (C57BL/10 SnJ) and *mdx* (C57BL/10 ScSn-*mdx*) mice obtained from Jackson Laboratories were maintained in barrier isolation facilities at the University of Michigan. The procedures used in this study were approved by the University of Michigan Committee on the Use and Care of Animals.

Measurements of cardiac myocyte length, tension and $[Ca^{2+}]_i$. Methods for isolating mouse adult cardiac myocytes were as previously reported²¹. Isolated single myocytes were transferred to an experimental chamber containing a silicon-coated glass bottom and platinum electrodes mounted on the sides to electrically stimulate the myocytes. The chamber was mounted on the stage of an inverted microscope (Nikon TE300, $\times 40$ objective) and kept at 37 °C with a thermoelectric device. The system for measuring length, tension and $[Ca^{2+}]_i$ in membrane-intact cardiac myocytes was developed from a previously reported method²², using a pair of microcarbon fibres. One fibre was connected to a sensitive force transducer system (200 mV mg⁻¹, Aurora Scientific) and the other was connected to a piezoelectric translator (P-173, PI Polytec) to control

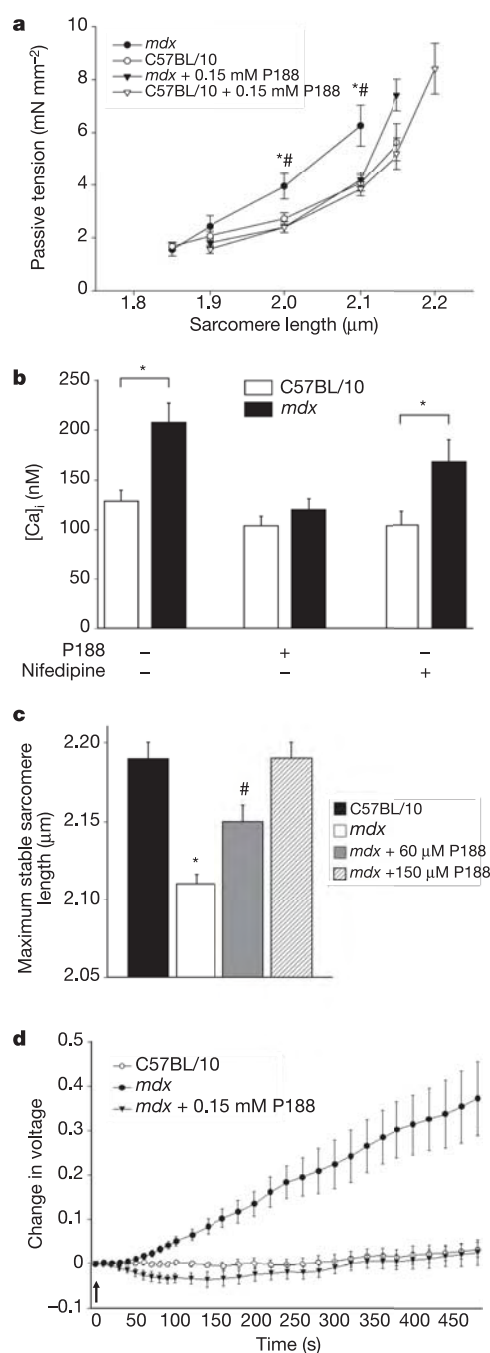


Figure 2 | Effects of P188 on passive tension and $[Ca^{2+}]_i$ during sarcomere length stretch in control and *mdx* single cardiac myocytes. **a, Effects of P188 on passive tension–extension relationships. *Mdx* myocytes have significantly higher passive tension than control (C57BL/10) myocytes (asterisk, $P = 0.038$ and $P = 0.031$ at sarcomere lengths 2.0 and 2.1 μm, respectively). Treatment with P188 significantly reduces tension compared with untreated *mdx* myocytes (hash symbol, $P = 0.021$ and $P = 0.040$ at sarcomere lengths 2.0 and 2.1 μm, respectively). **b**, Summary of the effects of P188 and nifedipine on $[Ca^{2+}]_i$ in myocytes upon sarcomere length stretch to 2.1 μm ($n = 7$). Asterisks indicate *mdx* greater than control myocytes ($P = 0.008$ in untreated myocytes, $P = 0.32$ in P188-treated myocytes and $P = 0.044$ in nifedipine-treated myocytes). **c**, Maximum tolerated sarcomere length stretch in *mdx* myocytes. Maximum tolerated sarcomere length is defined as the maximum length at which the myocyte is stable and fibrillations (Supplementary Video 3) are absent. Asterisk indicates *mdx* less than all groups ($P < 0.01$); hash denotes *mdx* + 60 μM P188 different from all groups ($P < 0.05$). Number of myocytes, $n = 7$ (control), 7 (*mdx*), 8 (*mdx* + 60 μM P188), 7 (*mdx* + 150 μM P188). **d**, FM 1-43 fluorescence after a single stretch in single cardiac myocytes. Summary of FM 1-43 fluorescence over time after a 20% stretch (arrow), showing mean change in voltage from baseline at the beginning of the stretch; number of myocytes, $n = 7$ (control), 6 (*mdx*), 7 (*mdx* + 150 μM P188). For all graphs, data represent mean $\pm$ s.e.m.**

myocyte length. The carbon fibres used in this study were rigid (40- μm diameter; 0.02 m N^{-1} compliance) and not altered by the active and passive tensions produced by single cardiac myocytes. Tension signal from the force transducer and length signal applied to the piezoelectric translator were recorded at 1,000-Hz sampling rate with an analogue-to-digital recording system (Accura 100, Nicolet). Video recordings of the cardiac myocytes were digitized to measure cell extension and sarcomere length via images on a computer screen. For measurement of $[\text{Ca}^{2+}]_i$, myocytes were incubated with $5\text{ }\mu\text{M}$ Fura-2-acetoxymethyl (AM) ester (Molecular Probes) and 0.02% Pluronic F127 (Molecular Probes) for 4 min at 37°C . Fura-2 excitation was sampled at 100 Hz using a microscope-based fluorescence spectrometer (Photon Technology International). $[\text{Ca}^{2+}]_i$ was determined using a previously established method²³ after subtracting background fluorescence intensity. We investigated the effect of P188 (150 μM in extracellular buffer unless otherwise stated)⁸ and nifedipine (1 μM) on passive tension–extension characteristics and $[\text{Ca}]_i$ of

control and *mdx* single cardiac myocytes. In another procedure, cardiac myocytes were stretched during the rising phase of twitch tension (lengthening contraction). Upon restoring sarcomere length, isometric twitch tension was again measured (post-stretch) and compared with pre-stretch twitch tension.

Membrane injury assay (FM1-43 study). Single cardiac myocytes were stretched twice with 20% extension (1-s duration) of resting cell length (this extension is less than the critical level of stretch that results in contracture of *mdx* cardiac myocytes) in HEPES-tyrode solution containing 1.8 mM Ca^{2+} and 2.5 μM FM1-43 dye (Molecular Probes). In the initial stretch we confirmed the extension, and at the beginning of the second stretch ($t = 0$) we measured voltage output from a photomultiplier (PTi, Monmouth Junction) for 480 s, representing fluorescence intensity of a $15 \times 50\text{-}\mu\text{m}$ area on the myocytes.

Cardiac haemodynamics *in vivo*. Murine pressure–volume loops were acquired by previously described methods²⁴. Briefly, anaesthetized mice were ventilated with 2% isoflurane and 98% oxygen. A thoracotomy and pericardiotomy were performed, and a pressure–volume catheter was inserted into the left ventricle via an apical stab. Before catheter insertion, mice received an intravenous infusion of $\sim 150\text{ }\mu\text{l}$ of 10% human albumin with and without P188, at a rate of $\sim 200\text{ }\mu\text{l kg}^{-1}\text{ min}^{-1}$. After the collection of baseline haemodynamic data, dobutamine was infused at a dose of $42\text{ }\mu\text{g kg}^{-1}\text{ min}^{-1}$ (ref. 25). The dose of P188 (460 mg kg^{-1}) was shown to be effective at mitigating skeletal muscle damage following electrocution⁷. Acute cardiac failure was defined as a drop in left ventricular systolic pressure below 60 mm Hg, as below this level, rapid decompensation and death were frequent occurrences. Following the 30-min dobutamine challenge or the development of acute cardiac failure, volume measurements were calibrated as previously described²⁴.

Received 23 March; accepted 23 May 2005.
Published online 17 July 2005.

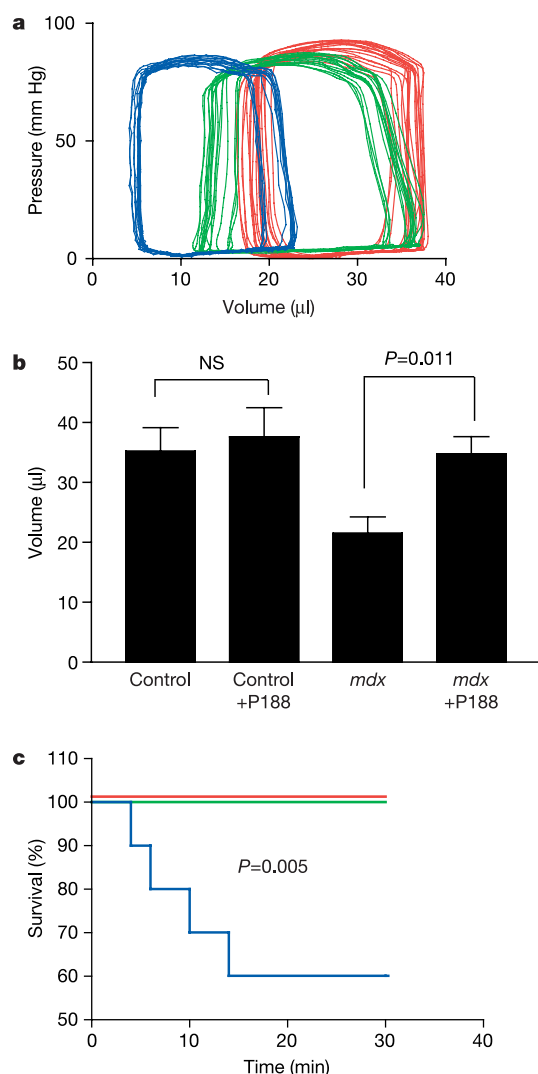


Figure 3 | Acute effects of P188 on *in vivo* haemodynamics and *mdx* survival. **a**, Representative pressure–volume loops in control (red) and *mdx* mice in the presence (green) or absence (blue) of acute infusion of P188. **b**, Summary of left ventricular end-diastolic volumes immediately following infusion of P188 in control and *mdx* mice. Values show mean $\pm$ s.e.m., $n = 13$ (control), 7 (control + P188), 13 (*mdx*), 11 (*mdx* + P188). NS, not significant. **c**, Kaplan–Meier survival analysis during infusion with $42\text{ }\mu\text{g kg}^{-1}\text{ min}^{-1}$ dobutamine. Control (red), *mdx* (blue), *mdx* + P188 (green). Mice were removed from the study when systolic pressures dropped below 60 mm Hg. Number of mice, $n = 13$ (control), 10 (*mdx*), 11 (*mdx* + P188).

- Muntoni, F. Cardiomyopathy in muscular dystrophies. *Curr. Opin. Neurol.* **16**, 577–583 (2003).
- Emery, A. E. H. in *Duchenne Muscular Dystrophy* (ed. Emery, A. E. H.) 26–45 (Oxford Univ. Press, Oxford, 2003).
- Finsterer, J. & Stollberger, C. The heart in human dystrophinopathies. *Cardiology* **99**, 1–19 (2003).
- Rodriguez, E. K. *et al.* A method to reconstruct myocardial sarcomere lengths and orientations at transmural sites in beating canine hearts. *Am. J. Physiol.* **263**, H293–H306 (1992).
- Pasternak, C., Wong, S. & Elson, E. L. Mechanical function of dystrophin in muscle cells. *J. Cell Biol.* **128**, 355–361 (1995).
- Wu, G. *et al.* Lipid corralling and poloxamer squeeze-out in membranes. *Phys. Rev. Lett.* **93**, 028101 (2004).
- Lee, R. C., River, L. P., Pan, F. S., Ji, L. & Wollmann, R. L. Surfactant-induced sealing of electroporabilized skeletal muscle membranes *in vivo*. *Proc. Natl Acad. Sci. USA* **89**, 4524–4528 (1992).
- Lee, R. C., Hannig, J., Matthews, K. L., Myerov, A. & Chen, C. T. Pharmaceutical therapies for sealing of permeabilized cell membranes in electrical injuries. *Ann. NY Acad. Sci.* **888**, 266–273 (1999).
- Turner, P. R., Fong, P. Y., Denetclaw, W. F. & Steinhardt, R. A. Increased calcium influx in dystrophic muscle. *J. Cell Biol.* **115**, 1701–1712 (1991).
- Togo, T., Alderton, J. M., Bi, G. Q. & Steinhardt, R. A. The mechanism of facilitated cell membrane resealing. *J. Cell Sci.* **112**, 719–731 (1999).
- Bansal, D. *et al.* Defective membrane repair in dysferlin-deficient muscular dystrophy. *Nature* **423**, 168–172 (2003).
- Danielou, G. *et al.* Dystrophin-deficient cardiomyocytes are abnormally vulnerable to mechanical stress-induced contractile failure and injury. *FASEB J.* **15**, 1655–1657 (2001).
- Davis, D. B., Delmonte, A. J., Ly, C. T. & McNally, E. M. Myoferlin, a candidate gene and potential modifier of muscular dystrophy. *Hum. Mol. Genet.* **9**, 217–226 (2000).
- Gregorevic, P. *et al.* Systemic delivery of genes to striated muscles using adeno-associated viral vectors. *Nature Med.* **10**, 828–834 (2004).
- Squire, S. *et al.* Prevention of pathology in *mdx* mice by expression of utrophin: analysis using an inducible transgenic expression system. *Hum. Mol. Genet.* **11**, 3333–3344 (2002).
- Torrente, Y. *et al.* Human circulating AC133⁺ stem cells restore dystrophin expression and ameliorate function in dystrophic skeletal muscle. *J. Clin. Invest.* **114**, 182–195 (2004).
- Goyenvalle, A. *et al.* Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. *Science* **306**, 1796–1799 (2004).
- Adams-Graves, P. *et al.* RheothRx (poloxamer 188) injection for the acute painful episode of sickle cell disease: a pilot study. *Blood* **90**, 2041–2046 (1997).
- Jewell, R. C., Khor, S. P., Kisor, D. F., LaCroix, K. A. & Wargin, W. A.

- Pharmacokinetics of RheothRx injection in healthy male volunteers. *J. Pharm. Sci.* **86**, 808–812 (1997).
20. Straub, V. & Campbell, K. P. Muscular dystrophies and the dystrophin-glycoprotein complex. *Curr. Opin. Neurol.* **10**, 168–175 (1997).
 21. Coutu, P., Bennett, C. N., Favre, E. G., Day, S. M. & Metzger, J. M. Parvalbumin corrects slowed relaxation in adult cardiac myocytes expressing hypertrophic cardiomyopathy-linked α -tropomyosin mutations. *Circ. Res.* **94**, 1235–1241 (2004).
 22. Yasuda, S. I. *et al.* A novel method to study contraction characteristics of a single cardiac myocyte using carbon fibers. *Am. J. Physiol. Heart Circ. Physiol.* **281**, H1442–H1446 (2001).
 23. Gryniewicz, G., Poenie, M. & Tsien, R. Y. A new generation of Ca^{2+} indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260**, 3440–3450 (1985).
 24. Michele, D. E., Gomez, C. A., Hong, K. E., Westfall, M. V. & Metzger, J. M. Cardiac dysfunction in hypertrophic cardiomyopathy mutant tropomyosin mice is transgene-dependent, hypertrophy-independent, and improved by β -blockade. *Circ. Res.* **91**, 255–262 (2002).
 25. Burger, A. J., Notarianni, M. P. & Aronson, D. Safety and efficacy of an accelerated dobutamine stress echocardiography protocol in the evaluation of coronary artery disease. *Am. J. Cardiol.* **86**, 825–829 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a grant from the National Institute on Aging (J.M.M.). D.T. was supported by an NIH National Research Service Award, and S.M.D. was supported by an American Heart Association Fellow-to-Faculty Transition Award.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M. (metzgerj@umich.edu).

LETTERS

ERM is required for transcriptional control of the spermatogonial stem cell niche

Chen Chen¹, Wenjun Ouyang⁴, Vadim Grigura¹, Qing Zhou⁵, Kay Carnes⁶, Hyunjung Lim³, Guang-Quan Zhao⁷, Silvia Arber^{8,9}, Natasza Kurpios¹⁰, Theresa L. Murphy¹, Alec M. Cheng⁴, John A. Hassell¹⁰, Varadaraj Chandrashekar¹¹, Marie-Claude Hofmann¹², Rex A. Hess⁶ & Kenneth M. Murphy^{1,2}

Division of spermatogonial stem cells¹ produces daughter cells that either maintain their stem cell identity or undergo differentiation to form mature sperm. The Sertoli cell, the only somatic cell within seminiferous tubules, provides the stem cell niche through physical support and expression of surface proteins and soluble factors^{2,3}. Here we show that the Ets related molecule⁴ (ERM) is expressed exclusively within Sertoli cells in the testis and is required for spermatogonial stem cell self-renewal. Mice with targeted disruption of *ERM* have a loss of maintenance of spermatogonial stem cell self-renewal without a block in normal spermatogenic differentiation and thus have progressive germ-cell depletion and a Sertoli-cell-only syndrome. Microarray analysis of primary Sertoli cells from ERM-deficient mice showed alterations in secreted factors known to regulate the haematopoietic stem cell niche. These results identify a new function for the Ets family transcription factors in spermatogenesis and provide an example of transcriptional control of a vertebrate stem cell niche.

Ets family transcription factors share a unique Ets DNA-binding domain and participate in a variety of developmental processes⁵. ERM^{4,6} belongs to a subfamily of Ets factors that also includes Pea3

and ER81 (ref. 5). Pea3 and ER81 are important for normal neuronal development^{7,8}. ERM is expressed in several tissues including brain, lung and testis⁶. To study the function of ERM *in vivo*, we generated mice with an inactivated *ERM* allele (*ERM*^{-/-}) (Supplementary Fig. 1). Heterozygous *ERM*^{+/-} mice are normal, and interbreeding heterozygous *ERM*^{+/-} mice yielded a mendelian (1:2:1) distribution of *ERM*^{-/-}, *ERM*^{+/-} and *ERM*^{+/+} in viable offspring (60:147:77, respectively), indicating that ERM is not critical for embryonic development or viability. The most marked phenotype in *ERM*^{-/-} mice is the disruption of spermatogenesis, whereas other organs did not reveal obvious anatomical defects (Supplementary Fig. 2).

Spermatogenesis is a cyclic process involving the differentiation of spermatogonial stem cells, meiotic cell division and the formation of haploid spermatids. Sertoli cells are the only somatic cells within seminiferous tubules and provide the immediate environment for developing germ cells². A balance between spermatogonial stem cell self-renewal and differentiation in the adult testis is essential for the maintenance of cyclic waves of spermatogenesis and fertility. Although *ERM*^{+/-} males were fertile, adult *ERM*^{-/-} males were sterile ($n = 12$). *ERM*^{-/-} males had a significantly decreased testicular size (Fig. 1a). At 4 weeks of age, seminiferous

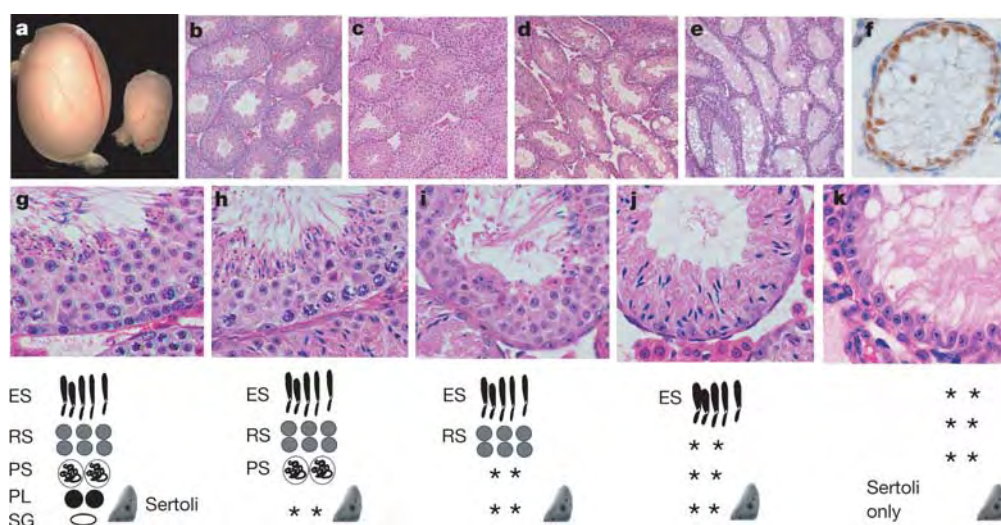


Figure 1 | Spermatogonial depletion and Sertoli-cell-only syndrome in *ERM*^{-/-} mice. **a**, Ten-week wild-type (left) and *ERM*^{-/-} (right) testes. **b**, Wild-type seminiferous tubules. **c–e**, Progressive germ-cell loss in *ERM*^{-/-} testis at 4 weeks (**c**), 6 weeks (**d**) and 10 weeks (**e**). **f**, Persistent Sertoli cells in *ERM*^{-/-} tubules indicated by GATA-1 expression at 10 weeks. **g–k**, Normal spermatogenesis with spermatogonial depletion. Histology of wild-type (**g**) or *ERM*^{-/-} (**h–k**) testes at 6 weeks of age. Diagrams of seminiferous epithelium are shown at the bottom: ES, elongated spermatid; RS, round spermatid; PS, pachytene spermatocyte; PL, preleptotene spermatocyte; Sg, spermatogonium; Sertoli, Sertoli cell; asterisks, missing cell populations.

¹Department of Pathology and Immunology, ²Howard Hughes Medical Institute, ³Department of Obstetrics and Gynecology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁴Department of Immunology, Genentech, 1 DNA Way, South San Francisco, California 94080. ⁵School of Molecular Biosciences, Washington State University, Pullman, Washington 99164, USA. ⁶Department of Veterinary Biosciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61802, USA. ⁷Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical School, Dallas, Texas 75390, USA. ⁸Biozentrum, Department of Cell Biology, University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland. ⁹Friedrich Miescher Institute, Maulbeerstrasse 66, 4058-Basel, Switzerland. ¹⁰Institute for Molecular Biology and Biotechnology, McMaster University, 1280 Main Street, Hamilton, Ontario L8S 4K1, Canada. ¹¹Department of Physiology, Southern Illinois University School of Medicine, Carbondale, Illinois, USA. ¹²Department of Biology, The University of Dayton, 300 College Park, Dayton, Ohio 45469-2320, USA.

tubules of wild-type and $ERM^{-/-}$ mice seemed similar, with multiple layers of germ cells indicating the normal initiation of spermatogenesis (Fig. 1b, c). By 6 weeks many $ERM^{-/-}$ tubules underwent progressive germ-cell maturation and depletion (Fig. 1d) and by 10 weeks most tubules were devoid of any germ cells (Fig. 1e), containing only morphologically normal Sertoli cells at the basement membrane, which expressed GATA-1 (Fig. 1f), a marker of mature Sertoli cells.

Histological analysis indicates that the primary defect in $ERM^{-/-}$ testis is the depletion of spermatogonia rather than a developmental block in spermatogenic differentiation. First, at 4 weeks or earlier, $ERM^{-/-}$ seminiferous tubules showed normal spermatogenic differentiation (Fig. 1c), indicating that $ERM^{-/-}$ primordial germ cells and embryonic gonocytes could give rise to spermatogonial stem cells and that a normal wave of spermatogenesis had occurred. However, by 6 weeks (Fig. 1g–k), an array of tubules showed either a selective loss of spermatogonia only or a combined loss of preleptotene and pachytene spermatocytes (Fig. 1h, i). Many tubules contained only elongated spermatids and Sertoli cells, with depletion of all precursor germ cells (Fig. 1j), whereas others were devoid of germ cells altogether (Fig. 1k). These data indicate that ERM deficiency is permissive of normal spermatogenic differentiation but causes germ-cell depletion through an initial loss of spermatogonial stem cells.

To assess the depletion of spermatogonia at the molecular level, we used microarray analysis to compare gene expression between

wild-type and $ERM^{-/-}$ testis at 4 weeks of age, when the testes of the two phenotypes have nearly identical histological appearances. Of 50 genes whose expression was reduced to less than one-third in $ERM^{-/-}$ testis compared with the wild type, the greatest reduction was observed for spermatogonia-specific genes (Fig. 2a and Supplementary Table 1). For example, the spermatogonia marker *Stra8* (ref. 9) was reduced 19-fold, and other spermatogonia-selective transcripts, including RNA-binding-motif protein (*Rbm*), *Dazl*, lymphoid-specific helicase (*Lsh*) and cellular retinoic-acid-binding protein (*CRABP*) were reduced 7-fold, 3.5-fold, 9-fold and 14-fold, respectively, in $ERM^{-/-}$ testis. By contrast, several markers of more mature germ cells were unchanged in expression, including the spermatid markers¹⁰ protamine 1 (*Prm1*) and *Prm2*, and transition proteins 1 and 2. Microarray results were confirmed by reverse-transcriptase-mediated polymerase chain reaction (RT-PCR) analysis in three distinct pairs of 4-week-old wild-type and $ERM^{-/-}$ littermates (Fig. 2a). Thus, at 4 weeks, when $ERM^{-/-}$ and wild-type testes are very similar in histological appearance, spermatogonia-specific genes were already greatly underrepresented in $ERM^{-/-}$ testes, again indicating that the defect is the loss of maintenance of spermatogonia. Finally, the loss of spermatogonial stem cells in $ERM^{-/-}$ mice was indicated by the absence of *Plzf1* expression in

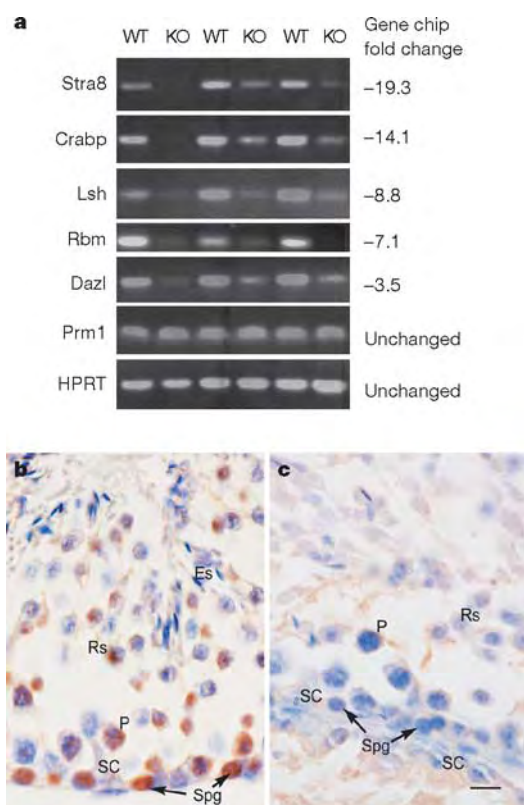


Figure 2 | Selective reduction of spermatogonia-specific genes in $ERM^{-/-}$ testes. **a**, RT-PCR analyses of three independent pairs of wild-type (WT) and $ERM^{-/-}$ (KO) male littermates are shown for the indicated genes. Numbers at the right represent the relative fold reduction measured by microarray analysis (Supplementary Table 1). HPRT, hypoxanthine-guanine phosphoribosyltransferase. **b**, **c**, Expression of *Plzf* in 6-week-old wild-type (**b**) and $ERM^{-/-}$ (**c**) testes by immunohistochemistry. Cells indicated are as follows: ES, elongated spermatid; RS, round spermatid; P, pachytene spermatocyte; Spg, spermatogonium; SC, Sertoli cell. Scale bar, 10 μ m.

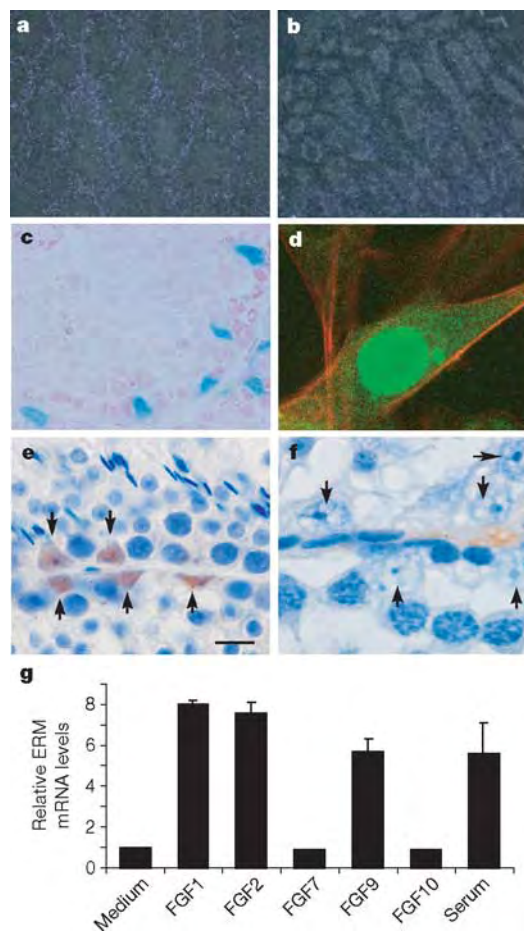


Figure 3 | ERM expression in testis is restricted to Sertoli cells. **a**, **b**, *In situ* hybridization for ERM mRNA in 10-week wild-type (**a**) and $ERM^{-/-}$ (**b**) testes. **c**, Testes from 6-week $ERM^{IRES-LacZ}$ heterozygous males were stained for LacZ expression. **d**, TM4 Sertoli cells were infected with ERM-GFP-RV retrovirus and analysed by confocal microscopy. **e**, **f**, Immunolocalization of ERM protein in adult wild-type (**e**) and $ERM^{-/-}$ (**f**) testes with anti-ERM monoclonal antibody. Arrows indicate nuclei of Sertoli cells in wild-type (**e**) and $ERM^{-/-}$ (**f**) testes. **g**, ERM mRNA expression in TM4 cells after treatment with various FGFs. Data are fold induction (means $\pm$ s.d. for two independent measurements) compared with ERM expression in medium.

$ERM^{-/-}$ testis in comparison with wild-type testis (Fig. 2b, c). The rare cells with morphological similarity to spermatogonia that remained in $ERM^{-/-}$ testis at 6 weeks of age were negative for Plzf expression, indicating that these are committed spermatogonia rather than undifferentiated spermatogonia.

Failure to maintain spermatogonia in $ERM^{-/-}$ testis could result from either an intrinsic requirement for ERM in germ cells or cell-extrinsic requirement for ERM in Sertoli or other cells. We first analysed ERM expression in $c-kit^{W/W-v}$ testes (Supplementary Fig. 3a), which have a Sertoli-cell-only phenotype¹². ERM expression was increased in germ-cell-free $c-kit^{W/W-v}$ testis relative to wild-type testis, indicating that it is expressed by somatic cells. In addition, ERM was expressed in isolated Sertoli cells, but not in isolated spermatogonia, pachytene spermatocytes or round spermatids (Supplementary Fig. 3b), whereas Stra8 was expressed exclusively in spermatogonia, as expected. Further, we used several additional methods to show that ERM is exclusively expressed within Sertoli cell in the testis. First, by *in situ* hybridization, ERM messenger RNA

was localized to the periphery of seminiferous tubules in wild-type testis but was absent centrally (Fig. 3a). The non-functional $ERM^{-/-}$ mRNA transcript was detected in the Sertoli-cell-only $ERM^{-/-}$ testis at 10 weeks (Fig. 3b), indicating direct expression by Sertoli cells. Second, we examined ERM expression histologically with an IRES-LacZ reporter cassette targeted to the ERM locus (N. Kurpios, S. Arber, J.A. Hassell, unpublished observations). ERM expression in $ERM^{+/IRES-LacZ}$ testis was found exclusively in Sertoli cells, was first detectable between 3 and 4 weeks of age and persisted throughout adulthood (Fig. 3c and Supplementary Fig. 4). This onset of ERM expression precedes the timing of spermatogonial loss, which is consistent with a requirement for ERM in the adult stem cell niche in the testis. Third, a fusion protein of ERM and green fluorescent protein (GFP) was localized to the nucleus of TM4 Sertoli cells (Fig. 3d). Last, we generated an ERM-specific monoclonal antibody, 3H7 (Supplementary Fig. 5), which identified ERM protein expression to be present exclusively within Sertoli cell nuclei of wild-type testis (Fig. 3e) and to be undetectable in $ERM^{-/-}$ testis (Fig. 3f). Thus, ERM expression in testis is specific to Sertoli cells, indicating that the loss of spermatogonia in $ERM^{-/-}$ testes might not be due to a cell-intrinsic defect of spermatogonia but rather to alterations in the microenvironment provided by Sertoli cells.

Signalling by fibroblast growth factor (FGF) has been reported to regulate the expression of ERM in Zebrafish¹³, and FGF9 deficiency in mice causes a possible defect in Sertoli cell differentiation¹⁴, prompting us to examine the effect of FGF on ERM expression in Sertoli cells (Fig. 3g). ERM mRNA was induced by FGF1, FGF2 and FGF9, but not by FGF7 and FGF10, in the murine Sertoli cell line TM4 (Fig. 3g). These results indicate a potential activity of FGF receptor signalling in regulating ERM expression and Sertoli cell function *in vivo*.

Spermatogonial depletion could result from alterations in spermatogonial proliferation or apoptosis. We therefore examined cell proliferation in $ERM^{-/-}$ testes at 3 and 4 weeks by labelling *in vivo* with bromodeoxyuridine (BrdU) (Fig. 4a–d). At 3 weeks, before the loss of germ cells, BrdU incorporation by spermatogonia was normal in $ERM^{-/-}$ testis (Fig. 4a, b). However, at 4 weeks, BrdU incorporation was almost absent in $ERM^{-/-}$ testis, compared with labelling in wild-type (Fig. 4c, d). To determine whether this loss was due to increased apoptosis, we used TdT-mediated dUTP nick end labelling (TUNEL), which showed no alteration in TUNEL staining in $ERM^{-/-}$ testes at 4 or 6 weeks (Fig. 4e, f, and Supplementary Table 2). Further, expression of proapoptotic and antiapoptotic genes in the Bcl-2 family also showed no difference between $ERM^{-/-}$ and wild-type testes (Fig. 4g). These results indicate that the defect was caused by decreased self-renewal of spermatogonial stem cells. Finally, serum hormone levels were measured to determine whether the pituitary–testis axis had contributed to the phenotype. We found no significant difference in serum levels of testosterone or follicle-stimulating hormone between wild-type and $ERM^{-/-}$ mice (Supplementary Table 3). Together, these results indicate that ERM expression by Sertoli cells might be required for spermatogonial stem cell self-renewal, and that loss of spermatogonial stem cells in $ERM^{-/-}$ testis is due to a defect in the stem cell niche provided by ERM-expressing Sertoli cells rather than an endocrine disorder.

Several somatic signalling pathways regulating spermatogonial stem cell self-renewal have recently been discovered^{3,15,16}. Mutations in two genes, promyelocytic zinc-finger (Plzf)^{11,17} and glial-cell-derived neurotrophic factor (GDNF)³ result in a defect in mammalian spermatogonial stem cell self-renewal. By comparison, the self-renewal defect in $ERM^{-/-}$ testes is more severe and differs mechanistically from that caused by mutations in Plzf or GDNF. First, Plzf is a transcription factor expressed by spermatogonia^{11,17}, not by Sertoli cells, and the loss of Plzf causes a spermatogonia-intrinsic defect. Loss of spermatogonia in Plzf-deficient mice is not complete even at 8 months^{11,17}, indicating that spermatogonial

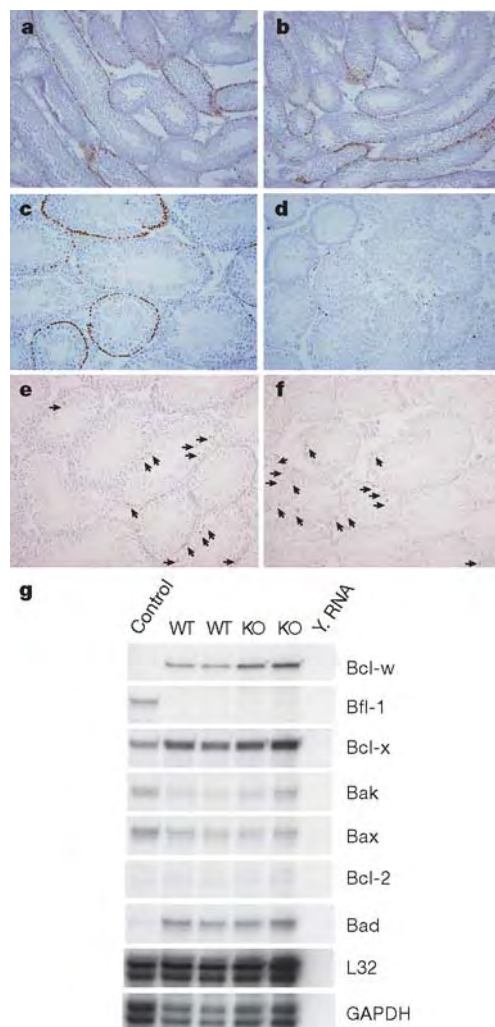


Figure 4 | Failure of stem cell self-renewal causes spermatogonial depletion. **a–d**, Loss of proliferating spermatogonia in $ERM^{-/-}$ testis, demonstrated by BrdU labelling *in vivo* of wild-type (**a**, **c**) and $ERM^{-/-}$ (**b**, **d**) testes at 3 weeks (**a**, **b**) and 4 weeks (**c**, **d**). **e**, **f**, Comparable levels of spermatogonial apoptosis, determined by TUNEL staining, in wild-type (**e**) and $ERM^{-/-}$ (**f**) testes at 4 weeks of age. Arrows indicate apoptotic cells. **g**, Expression of apoptosis-related genes was unchanged in $ERM^{-/-}$ testes as indicated by RNase protection assay of total RNA from 4-week-old wild-type (WT) and $ERM^{-/-}$ (KO) testes. GAPDH, glyceraldehyde3phosphate dehydrogenase. Y. RNA, Yeast tRNA.

self-renewal can occur for extended periods in the absence of Plzf, in contrast to more rapid and complete loss of spermatogonia in *ERM*^{-/-} testes. Second, GDNF³, which is secreted by Sertoli cells, was implicated in spermatogonial self renewal because GDNF^{+/-} mice show gradual spermatogonial depletion³. GDNF^{-/-} testes have not been analysed because of embryonic lethality.

To test whether ERM regulates GDNF, we performed a microarray analysis of purified wild-type and *ERM*^{-/-} primary Sertoli cells (Supplementary Fig. 6 and Supplementary Table 4). GDNF expression was unchanged in *ERM*^{-/-} testis in primary Sertoli cells from 4-week-old wild-type and *ERM*^{-/-} testes, indicating that it is not a target of ERM. In contrast, several other genes were greatly reduced in *ERM*^{-/-} Sertoli cells. Genes with the greatest reduction included the chemokines, CXCL-12 (SDF-1), CXCL5 (LIX) and CCL7 (MCP-3), reduced 9-fold, 10-fold and 25-fold, respectively (Supplementary Fig. 6 and Supplementary Table 4). SDF-1 and CXCL5 have been implicated in regulating the stem cell niche in other systems^{18,19}. SDF-1 is involved in haematopoietic stem cell (HSC) migration, retention and self-renewal¹⁸ and is required for the migration of primordial germ cells towards the genital ridge²⁰. CXCL5 was recently implicated in HSC maintenance¹⁹. Conceivably these chemokines are niche signalling molecules regulating spermatogonial stem cell self-renewal, but lethality caused by deficiencies of SDF-1 and its receptor CXCR4 prevent an immediate analysis of their role in adult spermatogenesis^{21,22}. Matrix metalloproteinase-12 (MMP-12) showed a tenfold reduction in *ERM*^{-/-} Sertoli cells. Interestingly, MMP-9, a related MMP family member, is involved in the recruitment of HSCs to the bone marrow niche²³.

Interactions with niche cells are crucial for maintaining stem cell character, and several molecules produced by niche cells can regulate the capacity of the niche to support stem cell self-renewal²⁴. As an example of the transcriptional control of a stem cell niche, ERM provides the opportunity to explore how the capacity of the stem cell niche to maintain spermatogonial self-renewal is transcriptionally coordinated.

METHODS

Generation of ERM mutant mice. ERM exons 2–5 encoding the initiation codon and transcriptional activation domain were deleted (Supplementary Fig. 1a). The targeting vector was constructed in pLNTK by using a 1.6-kilobase (kb) genomic fragment (left arm) upstream of the mouse ERM exon 2, and a 4-kb genomic fragment (right arm) downstream of exon 5. The left arm was generated by PCR from genomic DNA with the use of the oligonucleotides left arm forward (f), 5'-TTTGTGCGACGCGCGCTTTTGAATCTCTTAGGG AAGTTT-3' (*SalI* tailed), and left arm reverse (r), 5'-CCC CTCGAG TTTCCTCTGCTGTGTAGCCA-3' (*XhoI* tailed). The 1.6-kb PCR fragment was digested with *XhoI* and *SalI* and ligated into the *XhoI* site of pLNTK vector. The right arm was generated by PCR with the use of the oligonucleotides right arm forward (f), 5'-AAACTCGAGATACAAAGGATTGCAAAGGCT-3' (*XhoI* tailed), and right arm reverse (r), 5'-GGGACTCGAGTTCTGAAATTG TTTGGCCTTGA-3' (*XhoI* tailed), digested with *XhoI* and ligated into the *SalI* site of targeting vector. The targeting vector DNA was electroporated into MC50 embryonic stem cells (a gift from R. Schreiber). Positive clones were identified by Southern blot analysis with 5' and 3' probes (Supplementary Fig. 1a). *In vitro* Cre-mediated neo excision was performed on two distinct recombinant clones, 1CD3 and 1CC5, generating neo-deleted clones E7 and A7, respectively. Blastocyst injection was performed for all four clones and each generated germline transmission of the targeted ERM allele. Male chimaeras were crossed with 129SvEv females to establish ERM mutants on the 129SvEv genetic background. Homozygous mice were obtained by intercrossing heterozygous siblings. The phenotypes for all four lines were indistinguishable grossly and microscopically. For the results shown in this study, the E7 neo-deleted strain was used.

In situ hybridization. A 345-base-pair fragment of the ERM cDNA was obtained by RT-PCR with the use of the oligonucleotides ERM-345(f), 5'-CCGAGTT GTCGTCCTGTAG-3', and ERM-345(r), 5'-ACTGGCTTCAGGCATCATC-3', and cloned into pGEM-Teasy vector used for the synthesis of anti-sense and sense probes. Cryostat sections were hybridized with ³⁵S-labelled antisense RNA (cRNA) probe.

Generation of ERM-specific monoclonal antibody, and histology. ERM region

encoded by exons 7 and 8, lacking homology to Pea3 and ER81, was amplified by RT-PCR with the use of the primers 5'-GGAATTCATATGTGTGCTA CGATAGGAAGCCTCCC-3' and CGGGATCCTTATCTCTGTTCTGTATGGA TACTGG-3' and cloned into *NdeI/BamHI* sites of pET28a (Novagen). Histagged ERM recombinant protein (12 kDa) was induced with 1 mM isopropyl β-D-thiogalactoside in *Escherichia coli* BL21 (Invitrogen) and purified by Ni²⁺-nitrilotriacetate and size-exclusion chromatography. Hybridomas were generated from immunized hamsters and screened by ELISA against purified ERM protein. The hybridoma 3H7 monoclonal antibody (mAb) was used as supernatant for immunohistochemistry.

Immunohistochemistry was performed on sections fixed in 10% formalin. mAb 3H7 was used with goat anti-hamster biotinylated secondary antibody at 1:1000 dilution. Anti-GATA-1 rat mAb (Santa Cruz) was used at 1:100 dilution. Anti-Plzf antibody (Calbiochem) was used at 1:1000 dilution. Vectastain ABC kit and DAB substrate kit (Vector Laboratories) were used for immunohistochemistry. Sections were counterstained with haematoxylin. Analysis of the ERM-LacZ cassette used frozen sections stained overnight with 5-bromo-4-chloro-3-indolyl-β-D-galactoside staining buffer at 37°C, counterstained with nuclear Fast red.

An ERM-GFP fusion protein was created by deletion of the IRES from the plasmid ERM-RV²⁵ by Quick change mutagenesis (Stratagene) with the oligonucleotides ERM-GFP top (5'-CTTCGCTTACGTGAGCAAGGGCGAGGAGC-3' and ERM-GFP bot (5'-CCTTGCTCAGTAAGCGAAGCCTTCGGTGTGTA-3'), to produce ERM-GFP-RV. TM4 cells were infected with ERM-GFP-RV or GFP-RV and purified by cell sorting. Cellular localization of ERM-GFP fusion protein was captured by confocal microscopy.

TM4 cells were maintained in serum-free medium for 24 h and treated with medium alone or 1 nM FGF1, FGF2, FGF7, FGF9 and FGF10 (PeproTech) or 10% fetal calf serum. RNA was harvested after 3 h and real-time RT-PCR was performed for ERM and glyceraldehyde3phosphate dehydrogenase using an ABI Prism 7700 Sequence Detector (Applied Biosystems). ERM-specific primers used for this RT-PCR were 5'-CAGGAGCCCCGAGATTACTG-3' and 5'-CCGCTCTCATGTAGGATGAC-3'. Data are represented as fold expression of normalized ERM expression over the medium control.

Cell proliferation and apoptosis assays. Mice were injected intraperitoneally with BrdU (Sigma) at a concentration of 50 mg kg⁻¹. After 4 h, testes and small intestine were isolated and fixed in 10% buffered formalin. Paraffin-embedded sections were processed with the BrdU *in situ* detection kit (BD Biosciences). Germ-cell apoptosis was analysed by TUNEL labelling with Apoptag *in situ* apoptosis detection kit (Intergen) in accordance with the manufacturer's protocol. Expression of proapoptotic and antiapoptotic genes in 4-week-old wild-type and *ERM*^{-/-} testes was quantified by the mAPO-2 Multi-probe RNase protection assay kit (BD Biosciences) in accordance with the manufacturer's protocol.

Microarray analysis of wild-type and *ERM*^{-/-} testes and primary Sertoli cells. Wild-type and *ERM*^{-/-} testes (*n* = 4) were isolated at 4 weeks of age, and total RNA was extracted separately with RNeasy kit (Qiagen). RNA (10 μg) was pooled, and biotinylated cRNA target was independently generated from each pool. Each cRNA was hybridized to an Affymetrix U74Av2 Murine Genome Array. Microarray analysis of primary Sertoli cells isolated²⁶ from wild-type or *ERM*^{-/-} mice was carried out similarly.

RT-PCR analysis. We used semiquantitative PCR for confirmation of Affymetrix gene array results in Fig. 2a and Supplementary Fig. 5. RT-PCR amplifications were titrated within the linear range for each primer pair. Primers used are listed in Supplementary Table 5.

Received 8 March; accepted 6 June 2005.

- Kubota, H., Avarbock, M. R. & Brinster, R. L. Spermatogonial stem cells share some, but not all, phenotypic and functional characteristics with other stem cells. *Proc. Natl Acad. Sci. USA* **100**, 6487–6492 (2003).
- Griswold, M. D. The central role of Sertoli cells in spermatogenesis. *Semin. Cell Dev. Biol.* **9**, 411–416 (1998).
- Meng, X. *et al.* Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* **287**, 1489–1493 (2000).
- Monte, D., Baert, J. L., Defossez, P. A., de Launoit, Y. & Stehelin, D. Molecular cloning and characterization of human ERM, a new member of the Ets family closely related to mouse PEA3 and ER81 transcription factors. *Oncogene* **9**, 1397–1406 (1994).
- Seth, A. *et al.* The ets gene family. *Cell Growth Differ.* **3**, 327–334 (1992).
- Chotteau-Lelievre, A., Desbiens, X., Pelczar, H., Defossez, P. A. & de Launoit, Y. Differential expression patterns of the PEA3 group transcription factors through murine embryonic development. *Oncogene* **15**, 937–952 (1997).
- Arber, S., Ladle, D. R., Lin, J. H., Frank, E. & Jessell, T. M. ETS gene Er81 controls the formation of functional connections between group Ia sensory afferents and motor neurons. *Cell* **101**, 485–498 (2000).

8. Livet, J. *et al.* ETS gene *Pea3* controls the central position and terminal arborization of specific motor neuron pools. *Neuron* **35**, 877–892 (2002).
9. Oulad-Abdelghani, M. *et al.* Characterization of a premeiotic germ cell-specific cytoplasmic protein encoded by *Stra8*, a novel retinoic acid-responsive gene. *J. Cell Biol.* **135**, 469–477 (1996).
10. Sassone-Corsi, P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* **296**, 2176–2178 (2002).
11. Buaas, F. W. *et al.* *Plzf* is required in adult male germ cells for stem cell self-renewal. *Nature Genet.* **36**, 647–652 (2004).
12. De Franca, L. R. *et al.* Sertoli cells in testes containing or lacking germ cells: a comparative study of paracrine effects using the *W* (*c-kit*) gene mutant mouse model. *Anat. Rec.* **240**, 225–232 (1994).
13. Raible, F. & Brand, M. Tight transcriptional control of the ETS domain factors *Erm* and *Pea3* by *Fgf* signalling during early zebrafish development. *Mech. Dev.* **107**, 105–117 (2001).
14. Colvin, J. S., Green, R. P., Schmahl, J., Capel, B. & Ornitz, D. M. Male-to-female sex reversal in mice lacking fibroblast growth factor 9. *Cell* **104**, 875–889 (2001).
15. Kiger, A. A., White-Cooper, H. & Fuller, M. T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407**, 750–754 (2000).
16. Tran, J., Brenner, T. J. & DiNardo, S. Somatic control over the germline stem cell lineage during *Drosophila* spermatogenesis. *Nature* **407**, 754–757 (2000).
17. Costoya, J. A. *et al.* Essential role of *Plzf* in maintenance of spermatogonial stem cells. *Nature Genet.* **36**, 653–659 (2004).
18. Christensen, J. L., Wright, D. E., Wagers, A. J. & Weissman, I. L. Circulation and chemotaxis of fetal hematopoietic stem cells. *PLoS Biol.* **2**, E75 (2004).
19. Choong, M. L., Yong, Y. P., Tan, A. C., Luo, B. & Lodish, H. F. *LIX*: a chemokine with a role in hematopoietic stem cells maintenance. *Cytokine* **25**, 239–245 (2004).
20. Molyneux, K. A. *et al.* The chemokine *SDF1/CXCL12* and its receptor *CXCR4* regulate mouse germ cell migration and survival. *Development* **130**, 4279–4286 (2003).
21. Nagasawa, T. *et al.* Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine *PBSF/SDF-1*. *Nature* **382**, 635–638 (1996).
22. Tachibana, K. *et al.* The chemokine receptor *CXCR4* is essential for vascularization of the gastrointestinal tract. *Nature* **393**, 591–594 (1998).
23. Heissig, B. *et al.* Recruitment of stem and progenitor cells from the bone marrow niche requires *MMP-9* mediated release of *kit-ligand*. *Cell* **109**, 625–637 (2002).
24. Fuchs, E., Tumber, T. & Guasch, G. Socializing with the neighbors: stem cells and their niche. *Cell* **116**, 769–778 (2004).
25. Ouyang, W. *et al.* The *Ets* transcription factor *ERM* is Th1-specific and induced by *IL-12* through a *Stat4*-dependent pathway. *Proc. Natl Acad. Sci. USA* **96**, 3888–3893 (1999).
26. Karl, A. F. & Griswold, M. D. Sertoli cells of the testis: preparation of cell cultures and effects of retinoids. *Methods Enzymol.* **190**, 71–75 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. White for blastocyst microinjections; X. Cheng for helpful discussion; B. Sleckman for targeting vector; D. Ornitz, M. Griswold and K. Sheehan for reagents and antibody production; and T. Jessell for help with *ERM^{ires-LacZ}* mice. J.A.H. acknowledges funding from the Canadian Breast Cancer Research Alliance, the Canadian Institutes of Health Research and the DOD Breast Cancer Research Program. N.A.K. was supported by a DOD Breast Cancer Research Program Scholarship. K.M.M. is an investigator of the Howard Hughes Medical Institute.

Author Information Full Affymetrix data sets have been deposited with the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) as accession series GSE2205. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.M.M. (murphy@pathbox.wustl.edu).

Deficiency of *glutaredoxin 5* reveals Fe-S clusters are required for vertebrate haem synthesis

Rebecca A. Wingert¹, Jenna L. Galloway¹, Bruce Barut¹, Helen Foott¹, Paula Fraenkel¹, Jennifer L. Axe¹, Gerhard J. Weber¹, Kimberly Dooley¹, Alan J. Davidson¹, Bettina Schmidt¹, Barry H. Paw², George C. Shaw², Paul Kingsley³, James Palis³, Heidi Schubert⁴, Opal Chen⁴, Jerry Kaplan⁴, the Tübingen 2000 Screen Consortium‡ & Leonard I. Zon¹

Iron is required to produce haem and iron-sulphur (Fe-S) clusters, processes thought to occur independently^{1,2}. Here we show that the hypochromic anaemia in *shiraz* (*sir*) zebrafish mutants is caused by deficiency of *glutaredoxin 5* (*grx5*), a gene required in yeast for Fe-S cluster assembly. We found that *grx5* was expressed in erythroid cells of zebrafish and mice. Zebrafish *grx5* rescued the assembly of Δ *grx5* yeast Fe-S, showing that the biochemical function of *grx5* is evolutionarily conserved. In contrast to yeast, vertebrates use iron regulatory protein 1 (IRP1) to sense intracellular iron and regulate mRNA stability or the translation of iron metabolism genes^{1,2}. We found that loss of Fe-S cluster assembly in *sir* animals activated IRP1 and blocked haem biosynthesis catalysed by aminolaevulinic synthase 2 (ALAS2). Overexpression of ALAS2 RNA without the 5' iron response element that binds IRP1 rescued *sir* embryos, whereas overexpression of ALAS2 including the iron response element did not. Further, antisense knockdown of IRP1 restored *sir* embryo haemoglobin synthesis. These findings uncover a connection between haem biosynthesis and Fe-S clusters, indicating that haemoglobin production in the differentiating red cell is regulated through Fe-S cluster assembly.

sir^{h107} embryos had a blood-specific phenotype in which the circulating blood cells were hypochromic (or pale), and the mutation was lethal between 7 and 10 days post fertilization (d.p.f.). *sir* primitive erythrocytes have a defect in haemoglobin production (Fig. 1a). By β e1 *globin* mRNA expression, mutant embryos displayed a normal number of erythrocytes; however, a decrease in blood cell number was evident later (about 30% of wild-type cell number) (Fig. 1a). *sir* embryos continued to have hypochromia after 5 d.p.f., indicating that definitive haematopoiesis was defective (data not shown).

A chromosomal walk performed with a meiotic panel of 2,434 diploid *sir* embryos uncovered a deletion of about 150 kilobases on LG20 (Fig. 1b; data not shown). Using synteny of the locus to human chromosome 14q32, we identified several genes within the deletion. The zebrafish orthologue of *glutaredoxin 5* (*grx5*) was considered a likely candidate. Zebrafish *grx5* is 156 amino acid residues long and is highly conserved in eukaryotes (Fig. 1c). Glutaredoxins are thought to maintain the redox state, serving as essential protein antioxidants³. Studies in yeast found that *GRX5* was specifically required for the mitochondrial biogenesis of Fe-S clusters⁴⁻⁹, prosthetic groups that confer catalytic functions to proteins¹⁰. *GRX5*-deficient yeast lacked Fe-S enzyme activity, accumulated cellular iron and had a respiratory

deficit along with other symptoms of oxidative stress; these are phenotypes commonly associated with defects in Fe-S biogenesis^{4,11,12}.

grx5 was expressed in the intermediate cell mass, the zebrafish site of haematopoiesis, between 20 and 24 h.p.f. (Fig. 2a). *grx5* was also expressed in the liver and heart at 48 h.p.f. fertilization, with a low level of ubiquitous expression (Fig. 2a). *sir* mutants lacked *grx5* expression, which was consistent with the genomic deletion (Fig. 2a). *In situ* analysis of *grx5* expression during murine embryogenesis similarly found high expression in differentiating erythroblasts in the yolk sac and bloodstream (embryonic day (E)8.5–10.5) and definitive erythroblasts in the fetal liver (E12.5) (Fig. 2b). Northern blot analysis of adult murine tissues showed highest expression in bone marrow, spleen, heart, kidney and liver, although a low level was detected in all tissues (data not shown).

Antisense morpholino oligonucleotides (MOs) were used to assess the loss of *grx5* function during embryogenesis. MOs targeting the zebrafish *grx5* caused specific abrogation of haemoglobin synthesis when injected into wild-type embryos (MO1: *n* = 120 of 120 (100%); MO2: *n* = 148 of 148 (100%); Fig. 3a–c). This phenotype was rescued when *grx5* MOs were injected simultaneously with synthesized zebrafish *grx5* RNA (*n* = 297 of 314 (95%); Fig. 3d), and mismatch controls did not cause hypochromia (*n* = 94 of 94 (100%); data not shown). These findings provide strong evidence that *grx5* is responsible for the *sir* phenotype. We also overexpressed zebrafish *grx5* RNA in *sir* embryos, and observed rescue (*n* = 168 of 168 (100%); Fig. 3e). We next examined whether other orthologues could compensate for the absence of *grx5* in *sir*. Overexpression of yeast (*n* = 77 of 93 (83%)), mouse (*n* = 101 of 110 (92%)) or human (*n* = 84 of 96 (88%)) *grx5* RNA rescued *sir* hypochromia (Fig. 3f–h). These data suggest that the function of *grx5* is evolutionarily conserved.

In yeast, Grx5 is localized to the mitochondria, where it participates in Fe-S cluster assembly⁴⁻⁹. Consistent with a conserved function of *grx5* was our observation that green fluorescent protein (GFP) and Flag-tagged zebrafish Grx5 localized largely to the mitochondria (Supplementary Fig. 1; data not shown). Similarly, a proteomic survey of mouse mitochondria has found Grx5 in this organelle¹³. Some Grx5 in the cytoplasm indicates possible involvement with the cytoplasmic Fe-S machinery^{14,15}. We examined whether zebrafish *grx5* could rescue a yeast strain deficient in *GRX5*. Δ *grx5* yeast transformed with a plasmid carrying zebrafish *grx5* had wild-type activity of holoconitase, a cytoplasmic Fe-S-containing enzyme required for lysine synthesis, and fully restored activity of

¹Stem Cell Program and Division Hematology/Oncology Children's Hospital and Dana-Farber Cancer Institute, Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Division of Hematology, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Department of Pediatrics, University of Rochester Medical Center, Rochester, New York 14642, USA. ⁴Department of Pathology, School of Medicine University of Utah, Salt Lake City, Utah 84132, USA.

‡A list of members of the Tübingen 2000 Screen Consortium and their affiliations appears at the end of the paper.

the mitochondrial Fe–S enzyme aconitase (Fig. 3i, j). The ability of zebrafish *grx5* to substitute for yeast *GRX5* shows that *grx5* is probably required in vertebrates to make Fe–S clusters.

Loss of *grx5* function in yeast leads to mitochondrial iron accumulation and production of reactive oxygen species (ROS)^{4,11}. It is possible that the anaemia observed in *sir* mutants results from a similar mechanism. We examined iron accumulation using Perl's DAB stain in *sir* mutants whose survival was rescued to 4 weeks of age (Supplementary Fig. 2) after *grx5* mRNA injection. No increase in iron was evident (data not shown). The zebrafish *sauternes* mutant, deficient in ALAS2, also does not form iron-laden sideroblasts¹⁶, illustrating a difference between fish and human. To assay ROS levels, we subjected zebrafish to the indicator 5,6-chloromethyl-2',7'-dichlorodihydrofluorescein diacetate (CM-H₂DCFDA)¹⁷, and undertook flow cytometry analysis. In contrast to H₂O₂ treatment,

which induced ROS activity, there was no difference between wild-type and mutant *sir* embryos (data not shown). This suggested that an alternative mechanism was responsible for the *sir* hypochromic phenotype.

There has been no direct connection reported between Fe–S cluster biogenesis and haemoglobin production in erythroid cells. Unlike yeast, higher eukaryotes use iron regulatory proteins 1 and 2 (IRP1/2) to control cellular iron homeostasis through post-transcriptional regulation of iron uptake, storage and utilization genes^{1,2}. IRPs bind iron response elements (IREs), stem-loop structures located in the 5' or 3' untranslated region (UTR) of target RNAs, respectively blocking translation or stabilizing the transcript. The IRE-binding activity of IRP1 is negatively regulated by its interaction with a 4Fe–4S cluster¹⁸. When cellular iron is replete, IRP1 binds a 4Fe–4S cluster (converting it to a cytosolic aconitase), abolishing its IRE-binding activity. IRP2 function is regulated by means of proteasomal degradation when cellular iron is adequate, through binding to iron and possibly haem. We predicted that a block in Fe–S assembly in *sir* caused more active IRP1 to be present, inappropriately reflecting a low-iron state and

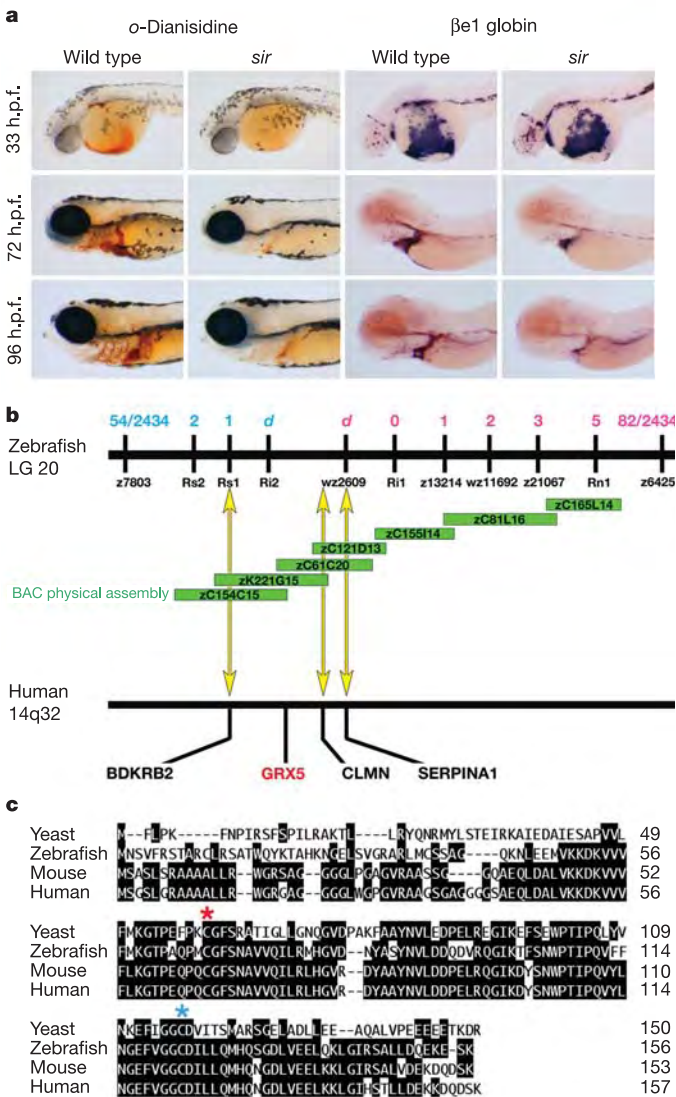


Figure 1 | *sir* characterization and positional cloning. **a**, o-Dianisidine staining for haemoglobin and β e1 globin whole-mount *in situ* hybridization for erythrocyte cell number, in wild-type and *sir* mutant embryos at 33, 72 and 96 h.p.f. Lateral views, anterior on left. **b**, Top: zebrafish linkage group 20, the number of meiotic recombination events given above genetic markers. BAC, bacterial artificial chromosome. Bottom: synteny to human 14q32 predicted that *grx5* was deleted in *sir*. **c**, Sequence alignment of yeast, zebrafish, mouse and human *grx5*, with residues identical to the consensus (black); orthologues show conservation of the active cysteine residues^{8,9} (asterisks).

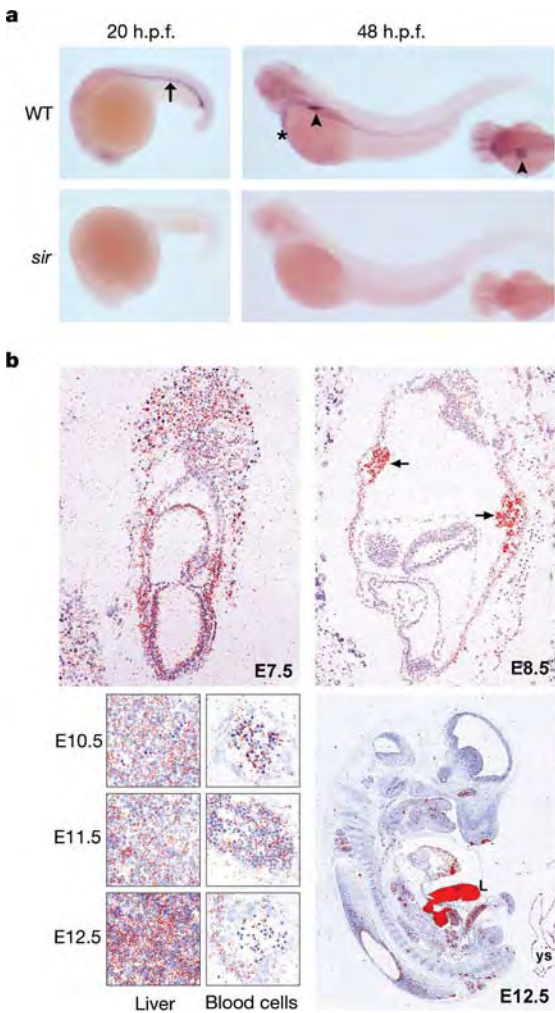


Figure 2 | Expression of *grx5* in vertebrates. **a**, Whole-mount *in situ* hybridization for *grx5* expression (stained purple) shows high expression in developing blood (arrow) at 20 h.p.f. and in the liver (arrowhead) and heart (asterisk) at 48 h.p.f. in wild-type zebrafish. *sir* mutants lack *grx5* expression. Lateral view, anterior on left; inset shows dorsal view. WT, wild type. **b**, Sagittal sections of mouse embryo *in situ* hybridization for *grx5* (stained red) show ubiquitous expression at E7.5, progressive downregulation of *grx5* in maturing primitive red cells between E10.5 and E12.5, and high expression in fetal liver at E12.5.

causing the depressed synthesis of some target genes. Because the target gene aminolaevulinic synthase 2 (*ALAS2*) encodes an erythroid-specific enzyme required for the first step of haem biosynthesis, and defects in this enzyme lead to hypochromic anaemia^{16,19}, a decrease in *ALAS2* protein amounts was a promising explanation for the hypochromic anaemia observed in *sir* embryos.

In this case, *sir* erythrocytes should be haem deficient. Whereas the injection of wild-type zebrafish embryos with MOs to the haem enzymes ferrochelatase (*FCH*) (Fig. 4a, b) or uroporphyrinogen decarboxylase (*UROD*) (data not shown) caused fluorescence or erythropoietic porphyria²⁰ ($n = 54$ of 54 and $n = 107$ of 107, respectively), *sir* embryos were unaffected ($n = 40$ of 40 and $n = 24$ of 24, respectively). This indicated an absence of haem intermediates in *sir* red cells and was consistent with an upstream block of *ALAS2*.

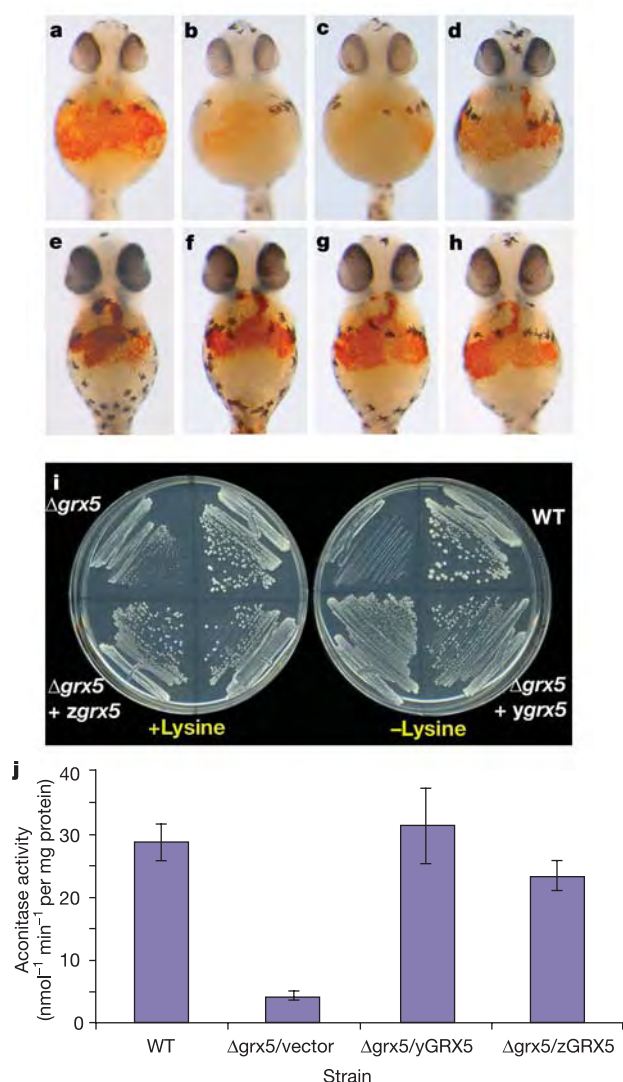


Figure 3 | Role of *grx5* in Fe-S biogenesis is conserved among eukaryotes. *o*-Dianisidine staining in embryos 36 h.p.f. **a**, Wild-type embryo; **b**, *sir* mutant; **c**, wild type injected with *grx5* MO; **d**, wild type injected with both *grx5* MO and *zgrx5* cDNA. **e–h**, *sir* mutants 40 h after fertilization injected with zebrafish *grx5* cDNA (**e**), yeast *grx5* cDNA (**f**), murine *grx5* cDNA (**g**) and human *grx5* cDNA (**h**). **i, j**, Overexpression of *zgrx5* in $\Delta grx5$ yeast: *zgrx5* caused rescue of lysine auxotrophy equivalent to *ygrx5* (**i**); measurement of aconitase activity in $\Delta grx5$ yeast shows comparable rescue of *zgrx5* and *ygrx5* (**j**). WT, wild type. Results are mean $\pm$ s.d. from four experiments, with significance $P < 0.001$ between all groups as determined by analysis of variance.

To evaluate whether an IRP1-mediated block of *ALAS2* translation caused *sir* haem deficiency, we overexpressed zebrafish *ALAS2* RNA without its 5' IRE. This rescued haemoglobin production in *sir* embryos ($n = 64$ of 81 (79%); Fig. 4c–e). Overexpression of native *ALAS2* RNA did not rescue *sir* ($n = 0$ of 66 (0%); Fig. 4f). A series of mutations shown to abrogate IRP1 or IRP2 IRE binding^{21,22} were made to the 5' IRE loop sequence (CAGUGC) of zebrafish *ALAS2*. Overexpression of *ALAS2* mRNA with a loop deletion ($n = 36$ of 51

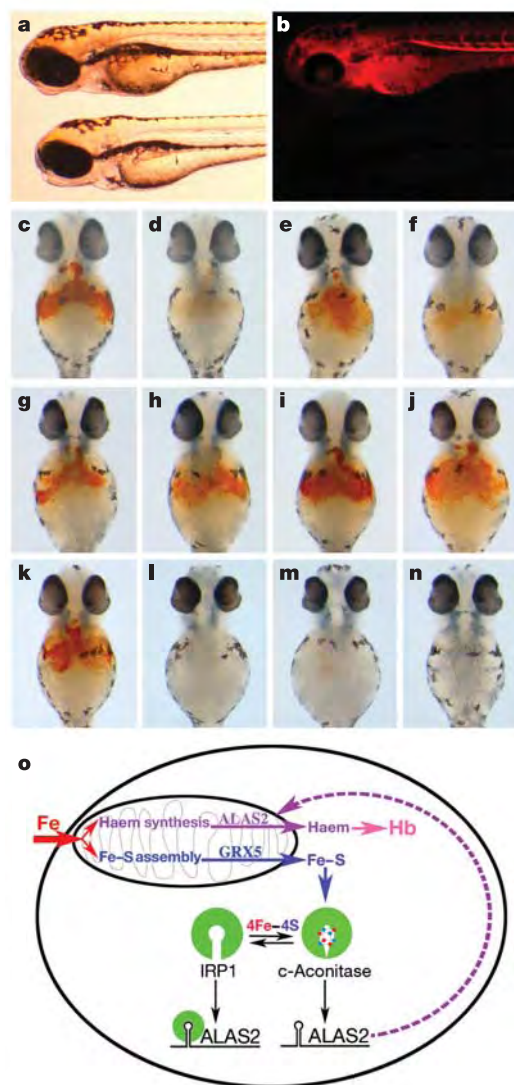


Figure 4 | Loss of Fe-S cluster production interferes with IRP1-mediated intracellular iron homeostasis. **a, b**, Wild-type embryos injected with *FCH* MO causes light-dependent erythrocyte autofluorescence (top), whereas *sir* (bottom) embryos are unaffected. **c–n**, *o*-Dianisidine staining in embryos 40 h.p.f. **c**, wild-type; **d**, *sir* mutant. **e–j**, *sir* mutants injected with *ALAS2* cDNA constructs: **e**, *ALAS2* cDNA lacking the 5' UTR; **f**, *ALAS2* cDNA with entire 5' UTR; **g**, *ALAS2* cDNA with 5' IRE six-base loop (CAGUGC) deleted; **h**, *ALAS2* cDNA loop 5'-GCTCGG-3'; **i**, *ALAS2* cDNA loop 5'-CCGAGC-3'; **j**, *ALAS2* cDNA loop 5'-CAGAGC-3'. **k–n**, *sir* mutants injected with *IRP1* MO (**k**) or *IRP1* four-base MO mismatch (**l**), injected with both *IRP1* MO and *IRP1* cDNA (**m**), and injected with *IRP2* MO (**n**). **o**, Model for the role of Fe-S cluster production in erythroid haem synthesis. Iron imported into mitochondria is used in the independent pathways of haem and Fe-S cluster biogenesis. Normal production of Fe-S clusters, which requires Grx5, switches the bifunctional protein IRP1 to a cytoplasmic aconitase (c-aconitase); this permits *ALAS2* protein synthesis needed for haem production. In the absence of Fe-S clusters, IRP1 has IRE-binding activity, and binds the 5' IRE on *ALAS2* mRNAs, blocking translation and thereby preventing haem production.

(71%), six-base mismatch (GCTCCG; $n = 40$ of 67 (60%)), two-base mismatch (CCGAGC; $n = 49$ of 67 (73%)) or single base deletions (CAGAGC; $n = 44$ of 51 (86%) and CAGUAC; $n = 43$ of 73 (59%)) rescued *sir* hypochromia (Fig. 4g–j; data not shown). These data supported the hypothesis that *sir* embryos were ALAS2 deficient because of a block in ALAS2 translation by IRP1.

Both IRP1 and IRP2 were found ubiquitously expressed from 8 to 72 h after fertilization during zebrafish embryogenesis, with high expression in the blood between 15 and 23 h.p.f. (Supplementary Fig. 3; data not shown). Antisense MOs were used to test whether IRP1 was responsible for the apparent shortage of ALAS2 in *sir*. Consistent with elevated IRP1 activity in *sir* embryos was our observation that MOs targeting *IRP1* restored *sir* haemoglobin synthesis ($n = 50$ of 95 (53%); $n = 33$ of 56 (60%); Fig. 4k), whereas mismatch controls did not ($n = 0$ of 104 (0%); $n = 61$ of 72 (85%); Fig. 4l, m). Although IRP2 is not regulated by Fe–S clusters, we tested whether IRP2 contributed to the ALAS2 block in *sir*. IRP2 MOs failed to rescue *sir* haemoglobin synthesis ($n = 0$ of 46 (0%); Fig. 4n). Polymerase chain reaction with reverse transcription (RT–PCR) confirmed that the IRP2 MOs abrogated normal RNA transcripts (data not shown). These findings indicate that IRP2 does not play a significant role in orchestrating changes to ALAS2 protein expression in *sir*.

Our studies of *sir* have identified a crucial role for mitochondrial Fe–S cluster assembly during haemoglobin production (Fig. 4o). Loss of Fe–S cluster biogenesis due to *grx5* deletion inappropriately activates IRP1, which blocks ALAS2 synthesis in *sir* embryos. This reveals a previously unknown link between the Fe–S and haem pathways. Mutations in the mitochondrial transporter ATP-binding cassette 7 (ABC7) cause X-linked sideroblastic anaemia and ataxia in humans^{23–25}; the yeast orthologue *ATM1* is deficient in cytosolic Fe–S proteins²⁶ and has been proposed to be the Fe–S cluster exporter¹². Our findings predict that ALAS2 deficiency caused by elevated IRP1 activity underlies ABC7-associated sideroblastic anaemia, and suggests that this unrecognized pathway is operative in human disease. Disruptions of Grx5 function in humans might similarly manifest as sideroblastic anaemia or might, as in the zebrafish, cause hypochromic, microcytic anaemia. Alternatively, Grx5 might be involved in iron overload disorders. The Grx5-null phenotype in yeast is suppressed by overexpression of *SSQ1* or *ISA2*, genes involved in Fe–S cluster synthesis⁴, which posits therapeutic targets for human diseases associated with perturbations in Grx5 activity. Our work on *sir* explains how complex biologic processes such as haemoglobin production in the red cell are directly regulated through Fe–S cluster assembly.

METHODS

Mutant characterization and genetic mapping. *sir*^{h107} embryos were generated on the Tü background, and zebrafish were maintained and staged as described²⁷. *o*-dianisidine staining²⁸, *in situ* hybridization¹⁶ of zebrafish embryos were performed as described. Linkage analysis was conducted as described²⁷ with diploid embryos obtained from *sir* Tü/WIK hybrids. Subsequent *sir* genotyping was performed with the forward and reverse primers 5'-TGAATCAGTTGAC-TACCGGAAGA-3' and 5'-TTCTCCCCTATGATTGGTTGAAG-3', respectively, to detect the *grx5* deletion.

Isolation of *grx5* and zebrafish IRP complementary DNAs. Full-length zebrafish *grx5* was amplified from whole-embryo RNA from embryos 24 h.p.f. by using the forward and reverse primers 5'-CGAACTTAACTGCTTCAAAT-3' and 5'-TGCAATAGAAATGTAATCATC-3', respectively, on the basis of blast comparison of yeast *GRX5* with the available zebrafish genome sequence (Sanger Institute, Cambridge, UK), and subcloned into pGEM-T Easy vector (Promega) for *in situ* analysis. Blast and syntenic comparison were used to identify *mgrx5* and *hgrx5*, and full-length cDNAs were obtained from Open Biosystems. *ygrx5* was PCR amplified from *Saccharomyces cerevisiae* with the forward and reverse primers 5'-ATGTTTCTCCCAAAATTCATCCC-3' and 5'-TCAACGATC-TTTGGTTCTTCTTC-3', respectively. Zebrafish *IRP1* and *IRP2* were isolated by RT–PCR from RNA from embryos 24 h.p.f. and subcloned into p-GEMT-Easy.

Mouse *in situ* hybridization and northern blot analysis. *In situ* hybridization of mouse embryos was performed as described²⁹. A murine Universal northern blot (Seegene) was probed with full-length *mgrx5* cDNA in accordance with the

manufacturer's protocol.

Subcellular localization of zebrafish *grx5*. Zebrafish *grx5* was subcloned into pEGFP N1 (Clontech) to create a carboxy-terminal fusion with GFP or pCMV-Tag (Stratagene) to create the C-terminal fusion with Flag. Plasmid was transfected into HEK-293 cells with the ProFectin Mammalian Transfection System (Promega), in accordance with the manufacturer's instructions. Porin was detected with anti-porin ANT (H-188), rabbit polyclonal antibody (SC-11433; Santa Cruz); anti-rabbit secondary antibody conjugated with Texas red (Santa Cruz). Nuclear staining was detected with 4',6-diamidino-2-phenylindole (Molecular Probes).

Yeast *grx5* overexpression and aconitase assay. The zebrafish *grx5* sequence coding sequence lacking the putative mitochondrial targeting sequence (codons 41–156) and containing a C-terminal Flag epitope was cloned by PCR with the forward and reverse primers 5'-GCGGGATCCGGACAGAGAAGCTGGAGG-3' and 5'-TCCCCGCGGTACTTGTGTCATCGTCATCCTTGTAAATCCTTTGACTCCTTTTCTTGATC-3', respectively, and placed behind the yeast *GRX5* promoter and mitochondrial targeting sequence (codons 1–33), which was generated by PCR with the forward and reverse primers 5'-GCGTTCTCGAGGGTGCCATTATGACGACTG-3' and 5'-CGCGGATCCCTCTGTGCTCAAATACATCCG-3', respectively. The fusion construct was placed in the yeast expression vector pTF63. These primers lead to a BamHI site between the yeast and zebrafish sequences, resulting in a Gly-Ser insertion. In cells transformed with this construct, the hybrid protein was properly targeted to yeast mitochondria and cleaved to generate the zebrafish enzyme (zGrx5). Wild-type (BY4741) or *Δgrx5* yeast strains (MATA *his3Δ1*, *leu2Δ0*, *met15Δ0*, *ura3Δ0*, *grx5::kanMX4*) were transformed with either wild-type *GRX5* or zGrx5 to test the ability of the constructs to complement the lysine auxotrophy due to holoaconitase deficiency. Mitochondrial aconitase activity in both transformed and non-transformed cells was assayed as described³⁰.

cDNA overexpression constructs and MO designs. Full-length *zgrx5*, *ygrx5*, *mgrx5* and *hgrx5* cDNAs were subcloned into the pCS2+ vector. Full-length *alas2*-pCS2+ was described previously³. Site-directed mutagenesis of this construct was used to generate the *alas2* IRE mutant series. The *alas2* open reading frame (–IRE) was amplified from this construct by PCR and subcloned into pCS2+. All cRNAs were synthesized with SP6 mMessage mMachine (Ambion). For expression in zebrafish embryos, about 50 pg of *grx5* yeast, mouse or human orthologue synthetic cRNA or 500–600 pg of each *alas2* cRNA was injected into one-cell-stage zebrafish embryos from adult *sir* heterozygous incrosses.

All MOs were synthesized by Gene Tools, and 1 nl was injected into embryos at the one-cell or two-cell stage from adult *sir* heterozygous incrosses. *zgrx5* start site MO1 (CTGTTCGACCTAAAAACGCTATTTCAT; mismatch, CTGACGACGTAAAGAGGCTATACAT) was injected at 1.0 mM alone or with 50 pg of *zgrx5* cRNA; MO2 (CGCTATTCATTGTGCAAGTCTGCGC) was injected at 2.0 mM. The *FCH* start MO (CCCATATTCAGCATCAGAATGCCTG; mismatch, CACA GATTCAATC ATCAGACTGCCTG) was injected at 1 mM. The *UROD* start MO (GTCCTTATCCATCATGACCCGGCTTC; mismatch, GTACTTGTCCCAAT GACCTGCTTC) was injected at 0.135 mM. *IRP1* MO1 (CGGTGTGTGCA TAAGGGTTCGTCAT; mismatch, CGATGTGTGTATACGGGTTCCGCAT) was injected at 0.1 mM or with 500 pg of *zIRP1*. *IRP1* exon 1/2 splice donor and acceptor MOs (CAAAATGATCTTACCATATCTGGG; mismatch, CAATTAAT GATGTTACGATATGTGGG) and (GGAGAAGGGCAGCTGCTCTACAAC; mismatch, GGACAAGGCCAGCAGCTCTAGAAAC) were injected together at 0.25 mM each or with 500 pg of *zIRP1* cRNA. *IRP2* splice acceptor MOs (GGATTTTGAACACCACTGAGAAGAG) and (GCCTTTAAACCAACCTA CATGAGA) were injected together at 0.25 mM each.

For all cRNA and MO rescue injections, live embryos were evaluated at 36–40 h.p.f. for the presence of red erythrocytes in circulation. Embryos with a robust red colour were put into a wild-type category, embryos with clear blood were put into a non-rescue group, and any embryo with only a hint of red was also placed in the non-rescue group. All embryos were then stained with *o*-dianisidine, photographed and genotyped.

Received 13 March; accepted 8 June 2005.

1. Napier, I., Ponka, P. & Richardson, D. R. Iron trafficking in the mitochondrion: novel pathways revealed by disease. *Blood* **105**, 1867–1874 (2005).
2. Hentze, M. W., Muckenthaler, M. U. & Andrews, N. C. Balancing acts: molecular control of mammalian iron metabolism. *Cell* **117**, 285–297 (2004).
3. Fernandes, A. P. & Holmgren, A. Glutaredoxins: glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system. *Antioxid. Redox. Sig.* **6**, 63–74 (2004).
4. Rodriguez-Manzanique, M. T. et al. Grx5 is a mitochondrial glutaredoxin required for the activity of iron/sulfur enzymes. *Mol. Biol. Cell* **13**, 1109–1121 (2002).

5. Mühlenhoff, U. *et al.* Components involved in assembly and dislocation of iron-sulfur clusters on the scaffold protein Isu1p. *EMBO J.* **22**, 4815–4825 (2003).
6. Vilella, F. *et al.* Evolution and cellular function of monothiol glutaredoxins: involvement in iron-sulfur cluster assembly. *Comp. Funct. Genom.* **5**, 328–341 (2004).
7. Alves, R., Herrero, E. & Sorribas, A. Predictive reconstruction of the mitochondrial iron-sulfur cluster assembly metabolism. II. Role of glutaredoxin 5. *Proteins Struct. Funct. Bioinformatics* **57**, 481–492 (2004).
8. Belli, G. *et al.* Structure–function analysis of yeast Grx5 monothiol glutaredoxin defines essential amino acids for the function of the protein. *J. Biol. Chem.* **277**, 37590–37596 (2002).
9. Tamarit, J. *et al.* Biochemical characterization of yeast mitochondrial Grx5 monothiol glutaredoxin. *J. Biol. Chem.* **278**, 25745–25751 (2003).
10. Beinert, H., Holm, R. H. & Münck, E. Iron-sulfur clusters: nature's modular, multipurpose structures. *Science* **277**, 653–659 (1997).
11. Rodríguez-Manzanique, M. T. *et al.* Grx5 glutaredoxin plays a central role in protection against protein oxidative damage in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **19**, 8180–8190 (1999).
12. Lill, R. & Kispal, G. Maturation of cellular Fe-S proteins: an essential function of mitochondria. *Trends Biochem. Sci.* **25**, 352–356 (2000).
13. Mootha, V. K. *et al.* Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* **115**, 629–640 (2003).
14. Land, T. & Rouault, T. A. Targeting of a human iron–sulfur cluster assembly enzyme, nifs, to different subcellular compartments is regulated through alternative AUG utilization. *Mol. Cell* **2**, 807–815 (1998).
15. Tong, W. H. & Rouault, T. Distinct iron–sulfur cluster assembly complexes exist in the cytosol and mitochondria of human cells. *EMBO J.* **19**, 5692–5700 (2000).
16. Brownlie, A. *et al.* Positional cloning of the zebrafish *sauternes* gene: a model for congenital sideroblastic anemia. *Nature Genet.* **20**, 244–250 (1998).
17. Dernbach, E. *et al.* Antioxidative stress-associated genes in circulating progenitor cells: evidence for enhanced resistance against oxidative stress. *Blood* **104**, 3591–3597 (2004).
18. Rouault, T. A. & Klausner, R. D. Iron-sulfur clusters as biosensors of oxidants and iron. *Trends Biochem. Sci.* **21**, 174–177 (1996).
19. Nakajima, O. *et al.* Heme deficiency in erythroid lineage causes differentiation arrest and cytoplasmic iron overload. *EMBO J.* **18**, 6282–6289 (1999).
20. Gross, U., Hoffmann, G. F. & Doss, M. O. Erythropoietic and hepatic porphyrias. *J. Inherit. Metab. Dis.* **23**, 641–661 (2000).
21. Jaffrey, S. R., Haile, D. J., Klausner, R. D. & Harford, J. B. The interaction between the iron-responsive element binding protein and its cognate RNA is highly dependent upon both RNA sequence and structure. *Nucleic Acids Res.* **21**, 4627–4631 (1993).
22. Henderson, B. R., Menotti, E. & Kühn, L. C. Iron regulatory proteins 1 and 2 bind distinct sets of RNA target sequences. *J. Biol. Chem.* **271**, 4900–4908 (1996).
23. Allikmets, R. *et al.* Mutation of a putative mitochondrial iron transporter gene (ABC7) in X-linked sideroblastic anemia and ataxia (XLSA/A). *Hum. Mol. Genet.* **8**, 743–749 (1999).
24. Bekri, S. *et al.* Human ABC7 transporter: gene structure and mutation causing X-linked sideroblastic anemia with ataxia with disruption of cytosolic iron-sulfur protein maturation. *Blood* **96**, 3256–3264 (2000).
25. Maguire, A. *et al.* X-linked cerebellar ataxia and sideroblastic anemia associated with a missense mutation in the ABC7 gene predicting V411L. *Br. J. Haematol.* **115**, 910–917 (2001).
26. Kispal, G., Csere, P., Prohl, C. & Lill, R. The mitochondrial proteins Atm1p and Nfs1p are essential for biogenesis of cytosolic Fe/S proteins. *EMBO J.* **18**, 3981–3989 (1999).
27. Wingert, R. A. *et al.* The *chianti* zebrafish mutant provides a model for erythroid-specific disruption of *transferrin receptor 1*. *Development* **131**, 6225–6235 (2004).
28. Ransom, D. G. *et al.* Characterization of zebrafish mutants with defects in embryonic hematopoiesis. *Development* **124**, 311–319 (1996).
29. Palis, J. & Kingsley, P. D. Differential gene expression during early murine yolk sac development. *Mol. Reprod. Dev.* **42**, 19–27 (1995).
30. Chen, O. S., Hemenway, S. & Kaplan, J. Genetic analysis of iron citrate toxicity in yeast: implications for mammalian iron homeostasis. *Proc. Natl Acad. Sci. USA* **99**, 16922–16927 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all members of the Zon laboratory for their advice and support. R.A.W. is a Harvard University Fellow of the Albert J. Ryan Foundation. L.I.Z. is an Investigator of the Howard Hughes Medical Institute. This work was supported by HHMI and NIH grants.

Author Information The zebrafish *grx5*, murine *grx5*, human *grx5*, zebrafish *IRP1* and zebrafish *IRP2* sequences are deposited in GenBank under accession numbers DQ083329, DQ083330, DQ083331, DQ083332 and DQ083333, respectively. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.I.Z. (zon@enders.tch.harvard.edu).

The Tübingen 2000 Screen Consortium F. van Bebber¹, E. Busch-Nentwich¹, R. Dahm¹, H. G. Frohnhofer¹, H. Geiger¹, D. Gilmour¹, S. Holley¹, J. Hooge¹, D. Jülich¹, H. Knaut¹, F. Maderspacher¹, C. Neumann¹, T. Nicolson¹, C. Nüsslein-Volhard¹, H. Roehl¹, U. Schönberger¹, C. Seiler¹, C. Söllner¹, M. Sonawane¹, A. Wehner¹, C. Weiler¹, U. Hagner², E. Hennen², C. Kaps², A. Kirchner², T. I. Koblizek², U. Langheinrich², C. Metzger², R. Nordin², M. Pezzuti², K. Schlombs², J. deSantana-Stamm², T. Trowe², G. Vacun², A. Walker², C. Weiler²

¹Max-Planck-Institut für Entwicklungsbiologie, Spemannstrasse 35, 72076 Tübingen, Germany. ²Artemis Pharmaceuticals GmbH, 72076 Tübingen, Germany.

LETTERS

Animal virus replication and RNAi-mediated antiviral silencing in *Caenorhabditis elegans*

R. Lu^{1*}, M. Maduro^{2*}, F. Li^{1,3}, H. W. Li¹, G. Broitman-Maduro², W. X. Li¹ & S. W. Ding^{1,3}

The worm *Caenorhabditis elegans* is a model system for studying many aspects of biology, including host responses to bacterial pathogens^{1,2}, but it is not known to support replication of any virus. Plants and insects encode multiple Dicer enzymes that recognize distinct precursors of small RNAs and may act cooperatively^{3–7}. However, it is not known whether the single Dicer of worms and mammals is able to initiate the small RNA-guided RNA interference (RNAi) antiviral immunity as occurs in plants⁸ and insects⁹. Here we show complete replication of the Flock house virus (FHV) bipartite, plus-strand RNA genome in *C. elegans*. We show that FHV replication in *C. elegans* triggers potent antiviral silencing that requires RDE-1, an Argonaute protein^{10,11} essential for RNAi mediated by small interfering RNAs (siRNAs) but not by microRNAs. This immunity system is capable of rapid virus clearance in the absence of FHV B2 protein, which acts as a broad-spectrum RNAi inhibitor^{9,12} upstream of *rde-1* by targeting the siRNA precursor. This work establishes a *C. elegans* model for genetic studies of animal virus–host interactions and indicates that mammals might use a siRNA pathway as an antiviral response.

We chose animal nodavirus FHV to determine whether *C. elegans* supports virus replication, because FHV replicates in yeast, plant, insect and mammalian cells¹³. The first genome segment of FHV, RNA1, encodes the entire viral contribution to the viral RNA-dependent RNA polymerase (RdRP) and replicates autonomously in the absence of RNA2, which depends on RNA1 for replication¹³ (Fig. 1a). Worm strains were generated to carry either a chromosomally integrated FR1 (for FHV RNA1) or FR2 (for FHV RNA2) transgene, designed to yield transcripts in the soma after heat induction that are identical in sequence to FHV genomic RNAs 1 and 2, respectively (Fig. 1b). Northern blot hybridizations detected a high-level accumulation of FHV RNA1 in worms carrying the FR1-3 transgene after transcription induction (Fig. 1c, lane 3). FHV RNA2 accumulated to high levels in strain FR1-3/FR2 carrying both FR1-3 and FR2 combined by genetic crosses (Fig. 1c, lane 5) but was undetectable in worms carrying FR2 alone (Fig. 1c, lane 6). Thus, the abundant viral RNAs detected in transgenic worm strains resulted from active RNA replication, because in the absence of RNA replication the initial heat-inducible transcripts were below the limit of detection in FR2 worms two days after induction.

Both FR1-3 and FR1-3/FR2 worms also contained abundant RNA3 (Fig. 1c, lanes 3 and 5). RNA3 is a subgenomic RNA transcribed during RNA1 replication from an internal site of the complementary, replicative intermediate of RNA1, (–)-RNA1 (Fig. 1a), and is not required in the initiation of FHV infection unlike the genomic RNAs. RNA1 replication initiated from FR1-2 transcripts was inefficient and became detectable only in the presence of RNA2 replication (Fig. 1c, lanes 2 and 4), which directs the

expression of the pre-capsid protein (pre-CP) essential for virion packaging (Fig. 1a). Taken together, the detection of self-replication of RNA1 and *trans* replication of RNA2 as well as the production of a subgenomic RNA in these worm strains provide evidence that *C. elegans* supports complete replication of the FHV RNA genome.

To investigate a possible induction of antiviral silencing by FHV RNA replication in worms, we first examined whether FHV accumulation in *C. elegans* requires expression of the FHV-encoded B2 protein, an RNAi suppressor that is active across the animal and plant kingdoms^{9,12} and is not known to have a direct role in FHV RNA replication¹³. We created a derivative of the FR1-3 transgene, FR1-3ΔB2 (Fig. 1b), containing a point mutation that abolished the B2 open reading frame but had no effect on the out-of-frame overlapping viral RdRP open reading frame (Fig. 1b). In contrast to high levels of RNA1 and RNA3 in FR1-3 worms (Fig. 1c, lane 3), we

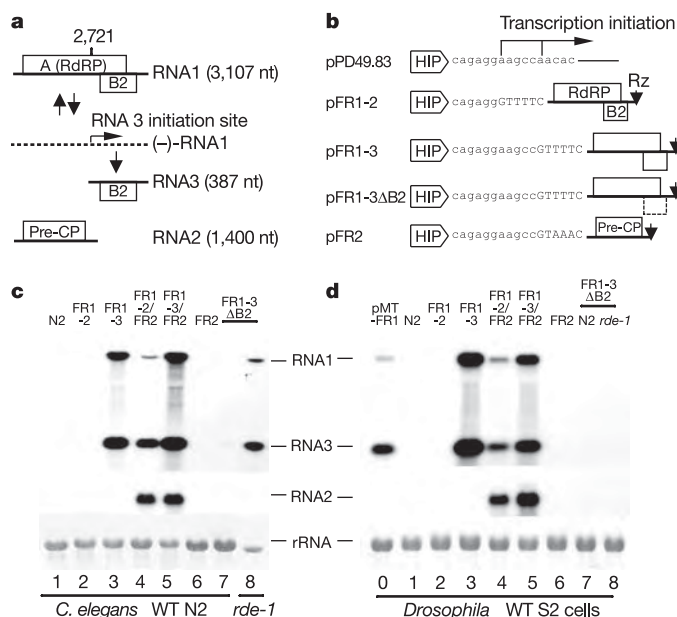


Figure 1 | Replication and silencing of FHV in *C. elegans*. **a**, Structure, replication and expression of FHV genome. **b**, Structure of FHV transgenes. HIV, heat inducible promoter. The junction region between HIV (lower-case letters) and FHV cDNA (capital letters) is shown. Transcriptional initiation sites from HIV were indicated. Rz is a self-cleaving ribozyme used to reduce non-viral extensions at the 3' termini of heat-induced transcripts. **c**, **d**, Northern blot detection of FHV RNA accumulation in *C. elegans* (**c**) and fruitfly S2 cells (**d**). Each lane in **c** and **d** was loaded with 4 μg of total RNA, except for lane 8 of **c**. WT, wild-type.

¹Institute for Integrative Genome Biology and Department of Plant Pathology, ²Department of Biology, and ³Graduate Program for Microbiology, University of California, Riverside, California 92521, USA.

*These authors contributed equally to this work.

detected no accumulation of FHV RNA1 in worms carrying an integrated FR1-3 Δ B2 transgene (Fig. 1c, lane 7). Thus, B2 is also essential for FHV accumulation in *C. elegans* as has been shown in cultured insect cells^{9,12}.

If the lack of FHV RNA accumulation in FR1-3 Δ B2 worms was indeed due to a loss of B2 suppression of RNAi, we would expect to find FR1-3 Δ B2 accumulation rescued in worm mutants defective in support of RNAi. To test this hypothesis, worm strains carrying FR1-3 and FR1-3 Δ B2 were each crossed into an *rde-1* (*ne219*) mutant background¹⁰. Unlike many RNAi mutants such as *dcr-1/Dicer*, which exhibit multiple developmental defects due to the disruption of microRNA (miRNA) function⁶, *rde-1* mutants are otherwise very healthy¹⁰. Northern blot analysis revealed that, whereas FHV RNA1 was undetectable in wild-type N2 worms carrying FR1-3 Δ B2 (Fig. 1c, lane 7), the same FR1-3 Δ B2 transgene directed abundant accumulation of FHV RNAs 1 and 3 in the *rde-1* mutant background (Fig. 1c, lane 8). In fact, the accumulation level of FHV RNAs in FR1-3 Δ B2/*rde-1* worms was similar to that in FR1-3 worms (Fig. 1c, compare lanes 3 and 8; note that the amount of total RNAs loaded in lane 8 was about one-third of that loaded in lane 3; see also Fig. 2). Thus, the mutant RNA1 from the FR1-3 Δ B2 transgene is not defective in self-directed replication, indicating that the reduced accumulation of viral RNA in FR1-3 Δ B2/N2 worms (Fig. 1c, lane 7) was due to the induction and clearance of viral RNAs by antiviral silencing in an *rde-1*-dependent siRNA pathway.

Both FHV RNAs 1 and 2 produced in worms were biologically active after transfection into cultured *Drosophila* S2 cells (Fig. 1d, lanes 3–5), which are known to initiate potent RNAi-mediated antiviral silencing after challenge with FHV⁹. In contrast, no viral RNA was detected in S2 cells transfected with total RNA extracted from either FR1-3 Δ B2/N2 or FR1-3 Δ B2/*rde-1* worms (Fig. 1d, lanes 7 and 8), in spite of the fact that the total RNA sample extracted from FR1-3 Δ B2/*rde-1* worms contained abundant FHV RNA1 (Fig. 1c, lane 8). This finding shows that progeny RNA1 of FR1-3 Δ B2 produced in FR1-3 Δ B2/*rde-1* worms remained defective in B2 expression and incapable of preventing virus clearance by RNAi-mediated antiviral silencing in S2 cells⁹. We therefore conclude that abundant accumulation of FHV RNAs detected in FR1-3 Δ B2/*rde-1* worms does not result from a genetic reversion in the progeny of FR1-3 Δ B2 but is due to the genetic suppression of viral B2 deletion by the loss-of-function mutation of *rde-1* in the worm genome. This genetic complementation of loss-of-function mutations in a viral RNAi suppressor gene and a host RNAi pathway gene provides the first direct evidence that viral RNAi suppressors enhance virus accumulation and facilitate viral infection by suppressing the host antiviral silencing mechanism.

Comparative analysis of the viral RNA accumulation among FR1-

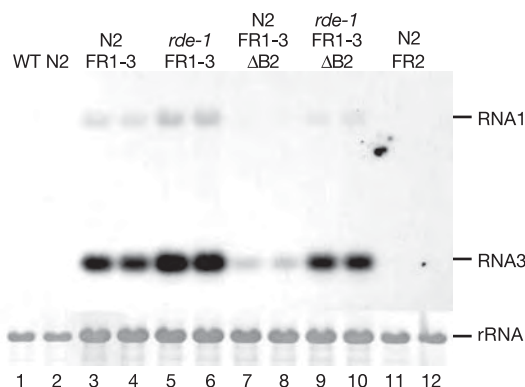


Figure 2 | FHV RNAi suppressor is active in *rde-1* worms. Total RNA was extracted from worms of either wild-type (WT) or *rde-1* genotype carrying an integrated FR1-3 or FR1-3 Δ B2 transgene, two days after transcriptional induction. Northern blot hybridizations were performed as in Fig. 1c, d.

3/*rde-1*, FR1-3/N2 and FR1-3 Δ B2/*rde-1* worms reveals two interesting results. First, the accumulation of FHV RNA1 derived from the same FR1-3 transgene array was much lower in N2 worms than in *rde-1* worms (Fig. 2, compare lanes 3 and 4 with lanes 5 and 6). This difference could be due to an incomplete B2 suppression of the worm *rde-1*-dependent antiviral silencing in FR1-3 worms, or alternatively it might result from the recently discovered, *rde-1*-dependent transcriptional transgene silencing in the soma¹⁴. Second, FHV RNA1 accumulated to much higher levels in FR1-3/*rde-1* worms than in FR1-3 Δ B2/*rde-1* worms (Fig. 2, compare lanes 5 and 6 with lanes 9 and 10), indicating that B2 expression also enhances viral accumulation in absence of *rde-1*. Because B2 is a broad-spectrum RNAi suppressor active in both the animal and plant kingdoms^{9,12} but does not influence the rate of FHV RNA replication¹³, our data indicate the possible presence of active antiviral RNAi in *rde-1* worms, which might be mediated by one or more of the remaining 26 worm Argonaute (AGO) genes, among which differential requirements for *alg-1*, *alg-2*, *ppw-1* and *ppw-2* in various RNAi processes have been documented¹⁵.

Our genetic analysis indicates that the viral RNAi suppressor is active in both wild-type and *rde-1* worms, indicating that B2 might act upstream of AGO. AGO has a key function in the RNA-induced silencing complex by binding to siRNA for recognition and cleavage of the mRNA target^{6,11}. No known protein domain was recognizable in the 106-residue FHV B2. We found that B2 synthesized as a glutathione *S*-transferase (GST) fusion was able to bind *in vitro* to a 21-nucleotide (nt) siRNA duplex (Fig. 3a, lanes 1–6) independently of its overhang nucleotides (data not shown); siRNA binding has been reported for p19, a silencing suppressor encoded by the plant tomodavirus¹⁶. However, unlike p19, which binds much more weakly to double-stranded RNA (dsRNA) longer than 23 nt, B2 also bound 25-nt siRNA and dsRNA 44 nt (Fig. 3d) and 100 nt (data not shown) in length. Furthermore, competition experiments indicated a much higher affinity of B2 for long dsRNA than for siRNA duplexes because the 100-nt dsRNA was approximately 30-fold more effective than 21-nt siRNA in inhibiting the formation of the siRNA–B2 complex (Fig. 3a, c).

A mutational analysis identified the replacement of Arg by Gln at position 54 (R54Q) of B2, which, when introduced into FHV RNA1, led to at least a 20-fold reduction in the accumulation of FHV RNAs in comparison with wild-type FHV RNA1 (data not shown). The same R54Q mutation completely abolished the B2 activity to bind long dsRNA (Fig. 3d, lanes 8–19). The effect of the R54Q mutation on siRNA binding was less marked: whereas the upper migrating siRNA–B2 complexes disappeared, the faster-migrating siRNA–B2 complexes remained visible (Fig. 3b, compare lanes 1, 2, 5 and 6 with lanes 3, 4, 7 and 8). Furthermore, we found that Dicer processing of a labelled 500-nt dsRNA into siRNAs was inhibited by FHV B2 fused with GST beginning at 800 nM (Fig. 3e, lane 9) but not by GST up to 10,000 nM (Fig. 3e, lane 6), and that the inhibitory effect was essentially eliminated by the R54Q mutation (data not shown). These findings indicate a new mechanism of viral suppression of antiviral silencing by targeting the dsRNA precursor of siRNAs, although they do not rule out a possible role of siRNA binding^{16–18}. This model is consistent with the genetic analysis placing B2 upstream of AGO and explains why B2 is active in both the animal and plant kingdoms.

Endogenous mammalian gene silencing by miRNAs starts in the nucleus with the cleavage by Drosha of primary miRNA transcripts into precursor miRNAs, which are then exported to the cytoplasm, where they are further processed into mature miRNAs by Dicer⁶. Much less is known about the initiation of antiviral RNA silencing in any organism. For RNA viruses, present models consider both dsRNA produced during RNA replication and highly structured elements in single-stranded viral RNAs as potential initiators by means of the siRNA and miRNA pathways, respectively^{16,18,19}. Our genetic analyses (ref. 9 and this study) indicate a possibly more

prominent role for the siRNA pathway in the recognition of the viral silencing initiators because neither *Drosophila* AGO2 nor *C. elegans* RDE-1 is essential for miRNA function²⁰, in spite of their requirement in antiviral silencing. Furthermore, with a few exceptions such as influenza viruses¹², and unlike DNA viruses²¹, most RNA viruses replicate exclusively in the cytoplasm and may escape detection by the miRNA pathway initiating in the nucleus. Nevertheless, potent antiviral silencing detected in a single Dicer organism rules out an essential role for multiple Dicers found in plants and insects^{3–7}, indicating an antiviral potential for the mammalian RNAi machinery through the siRNA pathway in addition to targeting viral mRNAs by cellular miRNAs²². The *C. elegans* model established in this work will facilitate genetic studies of animal virus–host interactions, many aspects of which cannot be addressed in the alternative model in *Saccharomyces cerevisiae*²³, a unicellular organism that does not seem to encode an RNAi pathway⁶. Although most nodaviruses are pathogens of insect and fish hosts, Nodamura virus¹² infects and

kills sucking mice and sucking hamsters¹³, indicating a potential of the worm model for studying the pathogenesis of mammalian viral diseases.

METHODS

Transgene constructs and transgenic worms. Constructs were generated by standard methods by using the vector pPD49.83 (a gift from A. Fire) which contains the promoter of the *hsp16-41* gene²⁴. Full-length cDNAs to FHV RNA1 and RNA2 together with a self-cleaving ribozyme from the tobacco ringspot virus satellite RNA were obtained from the infectious FHV cDNA clones previously constructed for infection of *Drosophila* S2 cells⁹.

Nematodes were propagated and maintained by standard protocols¹. The following strains were used: N2 (wild type), *unc-119(ed4) III* and *rde-1(ne219)* V. Animals were made transgenic by gonadal microinjection²⁵. FHV plasmids were mixed with either the *rol-6^D* plasmid pRF4 for injection into wild-type animals²⁵, or the *unc-119(+)* plasmid pDPMM016B for injection into *unc-119* mutants²⁶. Integrated lines were generated by treating about 30 transgenic hermaphrodites with 3,500 rad of γ -rays from a ¹³⁷Cs source, followed by screening for integrated animals in the F₂ generation.

rol-6^D and *unc-119(+)* transgenes were combined as described²⁶. To generate *rde-1;rol-6^D* strains, *rde-1(ne219)* males were crossed with *rol-6^D* transgene hermaphrodites. F₂ animals were grown individually on *Escherichia coli* HT115 expressing *unc-22* dsRNA (ref. 27), and plates generating no Unc-22 animals were assumed to be homozygous for *rde-1(ne219)*.

Assay for FHV replication in *C. elegans*. Expression of FHV transgenes was achieved for each strain as follows. First, ten young adult hermaphrodites were placed on a seeded 10-cm plate and allowed to grow to the next generation for 5 days at 20 °C. Plates were incubated for 2 h at 33 °C and returned to 23 °C for two days, after which they were washed off plates into water and stored at –80 °C. Total RNA was obtained by homogenizing thawed worm pellets with a Tissue Tearor (BioSpec Products) followed by extraction with TRIzol (Invitrogen), then a column-based RNA purification with the RNeasy kit (Qiagen) in accordance with the manufacturer's instructions. RNA concentrations were normalized and used for northern blot analysis in accordance with standard protocols⁹ with a ³²P-labelled cDNA probe that corresponded either to FHV RNA2 or to the last 387 nt of FHV RNA1, thus hybridizing to both RNAs 1 and 3.

To verify the biological activities of FHV RNAs produced in *C. elegans*, cultured *Drosophila* S2 cells were transfected with 2 μ g of total RNA extracted from individual worm strains two days after heat induction. Three days after transfection, total RNA was extracted from S2 cells for northern blot detection of FHV RNA accumulation. pMT-FR1 was used as a control and contained the full-length cDNA of FHV RNA1 under the transcriptional control of the CuSO₄-inducible metallothionein promoter as described⁹.

Fusion protein expression, dsRNA binding and Dicer activity assays. Expression of GST-tagged B2 proteins (GST-B2) in *E. coli* and RNA binding assays were performed as described previously^{12,28}. Both single-stranded and duplex siRNAs were synthesized chemically and end-labelled by exchange reaction. Long dsRNA (44 nt and 100 nt) was transcribed and annealed *in vitro*, either with or without kinase end-labelling after dephosphorylation. The final concentration for all labelled RNAs was 50 nM, whereas four different concentrations (50, 100, 500 and 5,000 nM) were used for each unlabelled competitor RNA. mB2 contained an Arg → Gln substitution at position 54 of FHV B2. The concentrations of GST-B2 or GST-mB2 used in RNA binding ranged from 0 to 40 ng μ l^{–1} as indicated. Binding reactions were resolved by native 6% polyacrylamide gel electrophoresis (PAGE) for binding to the 44-nt dsRNA or 8% for siRNA binding. Gels were dried before autoradiography.

Preparation of Dicer extracts from *Drosophila* S2 cells, labelling of long dsRNA, and the Dicer activity assay were as described⁷. A ³²P-labelled 500-nt dsRNA at 7.5 nM was incubated with the Dicer extracts in the presence of either GST or GST-B2 in five different concentrations, namely 50, 200, 800, 3,200 and 10,000 nM. After incubation, RNAs were fractionated by 15% PAGE along with a chemically synthesized 21-nt siRNA labelled with [γ -³²P]ATP as a marker.

Received 4 January; accepted 6 June 2005.

1. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
2. Mahajan-Miklos, S., Tan, M. W., Rahme, L. G. & Ausubel, F. M. Molecular mechanisms of bacterial virulence elucidated using a *Pseudomonas aeruginosa*–*Caenorhabditis elegans* pathogenesis model. *Cell* **96**, 47–56 (1999).
3. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
4. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, 642–652 (2004).
5. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).

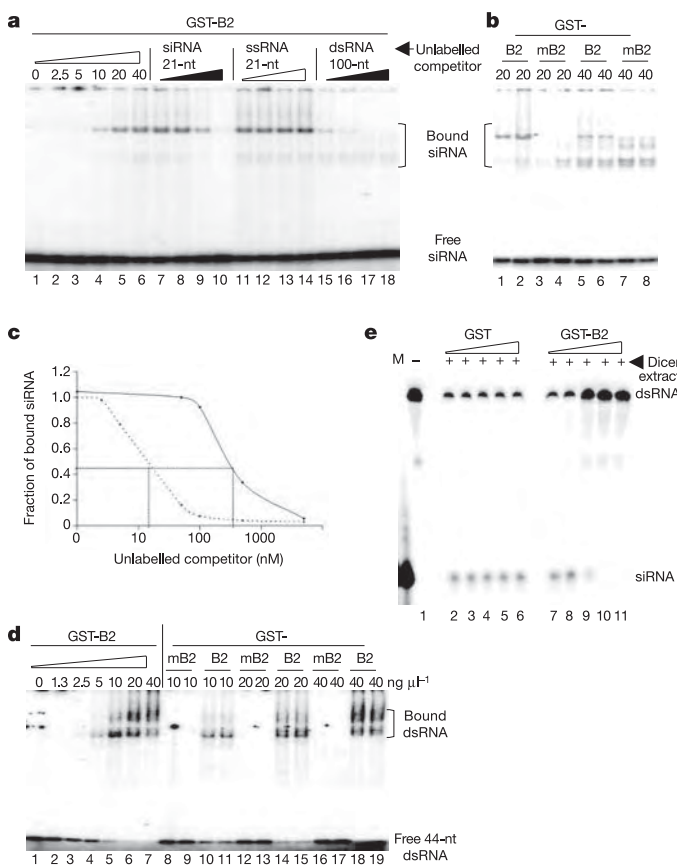


Figure 3 | FHV B2 is a dsRNA-binding protein and inhibits siRNA production *in vitro*. **a–d**, GST-tagged B2 protein (GST-B2) binds both 21-nt siRNA duplex (**a** and **b**) and 44-nt dsRNA (**d**) *in vitro*. Volumes of the top siRNA–B2 band in lanes 7–10 and 15–18 of **a** were quantified by phosphorimaging and plotted against the concentrations of the unlabelled competitor RNAs (**c**); dotted line, 100-nt dsRNA as unlabelled competitor; solid line, 21-nt siRNA as unlabelled competitor. Note the presence of non-specific signals in lanes 1, 8, 13 and 16 of **d**. The final concentration for all labelled RNAs was 50 nM whereas 4 different concentrations (50, 100, 500 and 5,000 nM) were used for each unlabelled competitor RNA (lanes 7–18 of **a**). The concentrations of GST-B2 or GST fused with the mutant B2 (GST-mB2) ranged from 0 to 40 ng μ l^{–1} in all RNA binding assays as indicated at the top of each lane, except for lanes 7–18 of **a** where GST-B2 was constant at 40 ng μ l^{–1}. **e**, B2 inhibits *in vitro* processing of a labelled long dsRNA by the Dicer extracts from fruitfly S2 cells. Concentration gradients of GST or GST-B2 used were 50, 200, 800, 3,200 and 10,000 nM. The labelled dsRNA was incubated in buffer without Dicer extract (lane –). A labelled 21-nt siRNA was used as a marker (M) at the left.

6. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529 (2005).
7. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
8. Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
9. Li, H. W., Li, W. X. & Ding, S. W. Induction and suppression of RNA silencing by an animal virus. *Science* **296**, 1319–1321 (2002).
10. Tabara, H. *et al.* The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
11. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
12. Li, W. X. *et al.* Interferon antagonist proteins of influenza and vaccinia viruses are suppressors of RNA silencing. *Proc. Natl Acad. Sci. USA* **101**, 1350–1355 (2004).
13. Ball, L. A. & Johnson, K. L. Reverse genetics of nodaviruses. *Adv. Virus Res.* **53**, 229–244 (1999).
14. Grishok, A., Sinskey, J. L. & Sharp, P. A. Transcriptional silencing of a transgene by RNAi in the soma of *C. elegans*. *Genes Dev.* **19**, 683–696 (2005).
15. Vastenhouw, N. L. & Plasterk, R. H. RNAi protects the *Caenorhabditis elegans* germline against transposition. *Trends Genet.* **20**, 314–319 (2004).
16. Silhavy, D. & Burgyn, J. Effects and side-effects of viral RNA silencing suppressors on short RNAs. *Trends Plant Sci.* **9**, 76–83 (2004).
17. Li, W. X. & Ding, S. W. Viral suppressors of RNA silencing. *Curr. Opin. Biotechnol.* **12**, 150–154 (2001).
18. Voinnet, O. Induction and suppression of RNA silencing: insights from viral infections. *Nature Rev. Genet.* **6**, 206–220 (2005).
19. Ding, S. W., Li, H., Lu, R., Li, F. & Li, W. X. RNA silencing: a conserved antiviral immunity of plants and animals. *Virus Res.* **102**, 109–115 (2004).
20. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
21. Pfeffer, S. *et al.* Identification of virus-encoded microRNAs. *Science* **304**, 734–736 (2004).
22. Lecellier, C. H. *et al.* A cellular microRNA directs antiviral immunity in human cells. *Science* **308**, 557–560 (2005).
23. Lee, W. M. & Ahlquist, P. Membrane synthesis, specific lipid requirements, and localized lipid composition changes associated with a positive-strand RNA virus RNA replication protein. *J. Virol.* **77**, 12819–12828 (2003).
24. Stringham, E. G., Dixon, D. K., Jones, D. & Candido, E. P. Temporal and spatial expression patterns of the small heat shock (hsp16) genes in transgenic *Caenorhabditis elegans*. *Mol. Biol. Cell* **3**, 221–233 (1992).
25. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
26. Maduro, M. F., Meneghini, M. D., Bowerman, B., Broitman-Maduro, G. & Rothman, J. H. Restriction of mesendoderm to a single blastomere by the combined action of SKN-1 and a GSK-3 β homolog is mediated by MED-1 and -2 in *C. elegans*. *Mol. Cell* **7**, 475–485 (2001).
27. Timmons, L., Court, D. L. & Fire, A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103–112 (2001).
28. Silhavy, D. *et al.* A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. *EMBO J.* **21**, 3070–3080 (2002).

Acknowledgements We thank X. Huang for recommending the use of pPD49.83, and the *Caenorhabditis* Genetics Center funded by the National Center for Research Resources of the National Institutes of Health for some of the strains used in this work. This project was supported by an NIH grant and USDA National Research Initiative Competitive Grants Program awards (to S.W.D.) and UC Riverside startup funds (to M.M.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.W.D. (dingsw@ucr.edu).

LETTERS

RNA interference is an antiviral defence mechanism in *Caenorhabditis elegans*

Courtney Wilkins¹, Ryan Dishongh¹, Steve C. Moore^{1,3}, Michael A. Whitt⁴, Marie Chow¹ & Khaled Machaca²

RNA interference (RNAi) is an evolutionarily conserved sequence-specific post-transcriptional gene silencing mechanism that is well defined genetically in *Caenorhabditis elegans*^{1–4}. RNAi has been postulated to function as an adaptive antiviral immune mechanism in the worm, but there is no experimental evidence for this. Part of the limitation is that there are no known natural viral pathogens of *C. elegans*. Here we describe an infection model in *C. elegans* using the mammalian pathogen vesicular stomatitis virus (VSV) to study the role of RNAi in antiviral immunity. VSV infection is potentiated in cells derived from RNAi-defective worm mutants (*rde-1*; *rde-4*), leading to the production of infectious progeny virus, and is inhibited in mutants with an enhanced RNAi response (*rrf-3*; *eri-1*). Because the RNAi response occurs in the absence of exogenously added VSV small interfering RNAs, these results show that RNAi is activated during VSV infection and that RNAi is a genuine antiviral immune defence mechanism in the worm.

Vesicular stomatitis virus (VSV), an enveloped virus with a non-segmented, negative-sense RNA genome, was an attractive candidate to infect *C. elegans* cells. VSV is particularly notable for its broad host range, including both mammalian and insect hosts⁵, and for its ability to infect and replicate at temperatures used in culturing *C. elegans* (18–25 °C). In addition, as the prototypical member of the Rhabdoviridae, it is among the best characterized of the negative-stranded RNA viruses, its life cycle has been extensively studied in mammalian cell culture⁶ and VSV recombinant viruses have been used as vectors for gene expression, vaccines and cell targeting^{7–10}.

A replication-competent, recombinant VSV (Fig. 1a) expressing the enhanced gene encoding green fluorescent protein (VSV-GFP) was able to infect primary cell cultures isolated from wild-type (N2) worms (Fig. 1b, VSV). GFP expression requires infectious virus because no fluorescent signals were detected in cells infected with virus that had been irradiated with ultraviolet (Fig. 1b, UV-VSV). In addition, antibodies against the VSV nucleocapsid (N) and glycoprotein (G) recognized only VSV-infected cells and not mock-infected cells (Fig. 1c). By using primers specific to N gene coding sequences in a reverse transcriptase-mediated polymerase chain reaction (RT-PCR) assay specific for positive-sense viral RNAs, a product was detected in VSV-infected cells (Fig. 1d, lane 2), but not in uninfected cells or cells infected with UV-VSV (Fig. 1d, lanes 1 and 3). The N gene RT-PCR signal, coupled with the presence of GFP and viral proteins (Fig. 1b, c), confirmed that synthesis of the viral transcripts was occurring within infected worm cells. Strand-specific RT-PCR products were also obtained from infected but not UV-VSV infected cells with the use of a primer set (P–M) amplifying the intergenic region between the P and M genes from negative-sense (genomic) viral RNAs (Fig. 1d). During genome replication, the viral polymerase synthesizes a full-length, positive-sense RNA (anti-

genome) species that is complementary to the viral genome and serves as a template for the synthesis of additional progeny RNA genomes. Because the intergenic regions are present in the full-length antigenome and progeny genome RNAs and in a small fraction of subgenomic, read-through messenger RNAs the P–M RT-PCR product from infected samples indicates that increased levels of viral genomes are present above that found in the initial inoculum of virus. Thus viral replication as well as viral gene expression was occurring in the infected *C. elegans* cells.

Establishment of a *C. elegans* model of VSV infection allowed us to test directly whether RNAi acts as an antiviral immune mechanism in the worm. RNAi is triggered by processing double-stranded RNAs

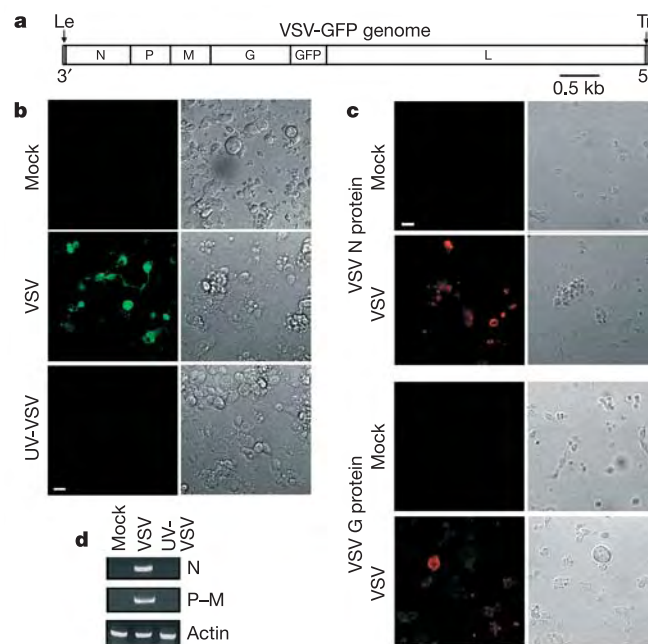


Figure 1 | VSV infection of *C. elegans* cells in culture. **a**, Genome organization of VSV-GFP recombinant virus. The structural proteins of the VSV virion are the nucleocapsid (N), phosphoprotein (P), matrix (M), glycoprotein (G) and viral replicase (L). **b–d**, Cells, mock infected or infected with VSV-GFP or UV-inactivated VSV-GFP (UV-VSV), were analysed at 48 h after infection. Scale bar, 7 μ m. **b**, GFP fluorescence images (left) and differential interference contrast images (right) of cells. **c**, Immunofluorescent staining of cells for VSV-N or VSV-G proteins. **d**, RT-PCR analyses for N gene sequences, the P to M intergenic region of the VSV genome and *C. elegans* actin.

¹Departments of Microbiology and Immunology and ²Physiology and Biophysics, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, USA. ³Department of Biology, Harding University, Searcy, Arkansas 72149, USA. ⁴Department of Molecular Sciences, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA.

(dsRNAs) into small interfering RNAs (siRNAs). These siRNAs associate with an RNA-induced silencing complex (RISC) to mediate the sequence-specific cleavage of target mRNAs and to induce post-transcriptional gene silencing (PTGS)^{2–4}. Genetic screens in the worm have identified several genes involved in this process^{11,12}; among them are *rde-1* and *rde-4* as potential modulators of viral immunity. RDE-1 is a member of the argonaute gene family¹¹ and RDE-4 is a dsRNA-binding protein that interacts with both the trigger dsRNA and RDE-1 (ref. 13). The RDE-4–RDE-1 complex functions at the initiation step of RNAi, where it is thought to recognize dsRNA and target it for cleavage into siRNAs by Dicer¹³. Null mutants in *rde-1* or *rde-4* are incapable of inducing RNAi but have no other apparent phenotypes¹¹. VSV infection of *rde-1* and *rde-4* mutant cells resulted in a higher percentage of GFP-positive cells in the culture (Fig. 2a, b). In addition, the mean GFP fluorescence intensities within individual cells continued to increase over time in infected *rde* cells (Fig. 2c). Because GFP is expressed from the fourth cytron, it is one of the least abundantly expressed proteins in VSV-GFP-infected cells, so the percentages of GFP-positive cells may be an underestimate of the total percentage of infected cells within the culture. To increase the sensitivity and obtain a more accurate

assessment, infected cells were also immunofluorescently stained for N protein (the most abundant viral protein in an infected cell) and analysed by flow cytometry. At 36 h after infection, the percentage of infected *rde-1* cells (61%) in the culture was again about double that for N2 cells (33%). Consistent with the increased numbers of infected cells, there were greater amounts of both N and G proteins immunoprecipitated from radiolabelled infected *rde* cell lysates than from N2 cell lysates (Fig. 2d).

To determine whether VSV infection is productive in *C. elegans*, N2 and *rde* cells were infected with VSV and the production of infectious virus was followed over time (Fig. 2e). A rapid and significant increase in viral titres was seen in infected *rde* cells between 24 and 30 h after infection, which reached a plateau by 36 h after infection (Fig. 2e). The kinetics of virus growth in the *rde* mutants is consistent with that normally observed during a synchronized single cycle of productive viral infection. In contrast, no significant increase in viral titres was detected at late times of infection in the culture supernatants of infected N2 cells. The small but continuous increase in virus titres observed over the entire infection period (0–48 h after infection) in N2 cells indicated that this titre might be due primarily to the elution of virus particles off

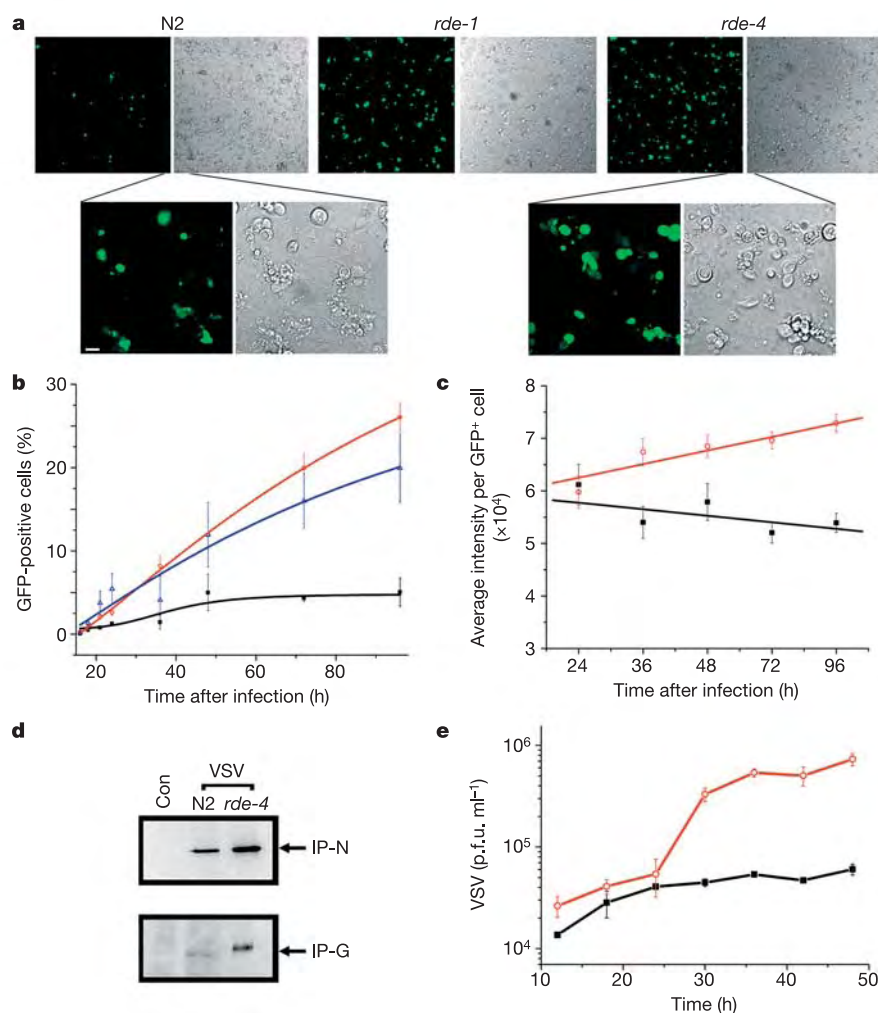


Figure 2 | Increased infection of RNAi-deficient mutants. **a–c**, GFP expression in VSV-GFP-infected N2, *rde-1* (ne219) and *rde-4* (ne301) cells. **a**, Confocal images at low magnification (top row) and high magnification (bottom row; scale bar, 7 μ m) at 48 h after infection. **b**, **c**, Percentage of GFP-positive cells (**b**) and intensities (four to eight individual fields) of GFP fluorescence per cell (**c**) at each time after infection. Black squares, N2; red

circles, *rde-4*; blue triangles, *rde-1*. Results are means $\pm$ s.e.m. **d**, Immunoprecipitation (IP) of VSV N and G proteins from equivalent numbers of infected N2 and *rde-4* cells. Con, control. **e**, Virus titres from infected N2 cells (black squares) and *rde-1* cells (red circles). Similar results were obtained with infected *rde-4* cells. Results are means $\pm$ s.e.m.

cells from the initial inoculum, and that no detectable amounts of new virus were produced in these cells.

Therefore, in an RNAi-null background, VSV infects a higher percentage of cells than in wild-type N2 cells, and both viral proteins and the GFP reporter accumulate to higher levels. Most significantly, VSV infections in *rde* mutants result in productive viral infections with significant increases in viral titres, which are absent from wild-type N2 cells. These data show that the RDE-1 and RDE-4 components of the RNAi pathway are important in inhibiting VSV infection, which limit the extent of viral gene expression and the production of infectious virus.

If an RNAi-null background potentiates viral infection, then mutations that enhance RNAi responses would be predicted to attenuate viral infection. Recently two worm mutants (*rrf-3* and *eri-1*) have been identified in which RNAi responses are enhanced by modulating intracellular siRNA levels^{14,15}. RRF-3, a member of the RNA-dependent RNA polymerase (RdRP) gene family in *C. elegans*, seems to inhibit RdRP-directed siRNA amplification, and worms with mutations in *rrf-3* are more sensitive, especially in neurons, to RNAi responses induced by dsRNAs¹⁴. ERI-1, a member of the DEDDh nuclease family, preferentially cleaves siRNAs. Thus, siRNAs are more stable and accumulate in *eri-1* mutants, resulting in enhanced gene suppression¹⁵. Comparison of VSV infection in *rrf-3* and *eri-1* mutants with that in N2 cells revealed that the percentages of VSV-infected cells were indeed lower in both *rrf-3* and *eri-1* than in N2 cultures (Fig. 3a). However, a small subpopulation of *eri-1* and *rrf-3* cells (about 2%) remained permissive to VSV infection, and GFP intensities in these VSV-permissive *rrf-3* and *eri-1* cells increased slightly over time (Fig. 3b), in contrast to the decrease observed in N2 cells. This indicates that the RNAi response might remain negatively

regulated in this subpopulation of *rrf-3* and *eri-1* cells, thus allowing the accumulation of viral proteins. A similar phenotype was observed in the whole worm in *eri-1* mutants, in which some neurons remain refractory to the RNAi response¹⁵. This virus-permissive phenotype could be explained either by the differential expression of the two known negative regulators of RNAi, *eri-1* and *rrf-3*, or perhaps by the presence of other as yet unidentified host suppressors of RNAi. Nevertheless, the decreased numbers of GFP-positive cells indicate that VSV infection is attenuated in host backgrounds displaying an enhanced RNAi response.

The phenotypes observed on infection of N2 and the RNAi-enhanced mutant cells indicate that the RNAi pathway might be induced on viral infection without the exogenous addition of synthetic siRNAs. RNase protection assays were performed to determine whether siRNAs were indeed generated as a consequence of viral infection (Fig. 4). Protected bands of about 20–30 nucleotides were observed when RNA from infected N2 cells was hybridized with radiolabelled VSV probes. In contrast, no protection was observed in reactions using RNA from mock-infected N2 or virus-infected *rde-4* cells. The presence of virus-specific siRNAs in infected N2 cells indicates that the RNAi response is induced on virus infection.

Collectively, the genetic and biochemical data in both RNAi-null and enhanced host backgrounds establish RNAi as a cellular antiviral defence mechanism in *C. elegans*. Pretreatment of mammalian cells with synthetic siRNAs against viral sequences, to initiate the RNAi pathway artificially, limits infection by several viruses including polio, influenza and HIV^{16–18}. This, coupled with the recognized role of RNAi in antiviral immunity in plants^{19–21} and insects^{22,23}, indicates that RNAi might serve a similar function throughout evolution. Moreover, we show here that, similarly to infections with plant or insect viruses, RNAi responses in the worm are induced as a consequence of VSV infection. These data, together with recent evidence that HIV infection can induce RNAi responses in human cells²⁴, indicate that activation of the RNAi pathway might be a routine physiological response of the host to viral infections. Indeed, many viruses, both plant and animal pathogens, have evolved suppressors of RNAi to bypass this cellular immunity and enable viral replication^{24–26}.

The relative genetic simplicity of *C. elegans* and the availability of an extensive collection of mutants makes the worm an attractive model host for identifying genetic determinants that affect the susceptibility or resistance of a host to viral infection. In addition to viral determinants, susceptibility to infection and disease is dependent on the genetic background and physiological status of

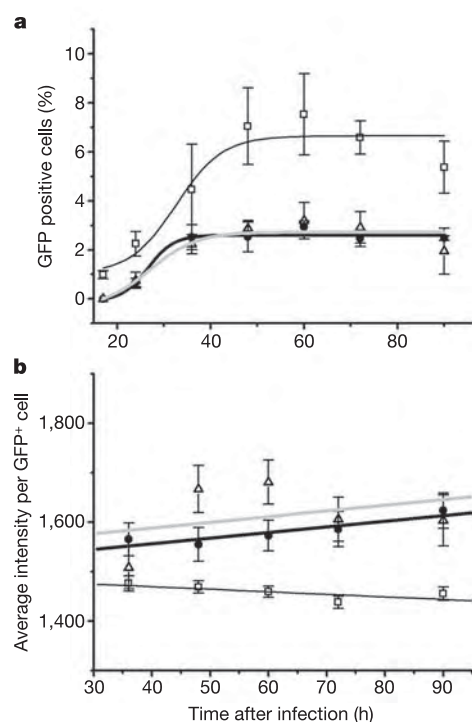


Figure 3 | Increased resistance of RNAi-hypersensitive mutants to VSV infection. Cells from N2 (squares), *eri-1*(*mg 336*) (filled circles) and *rrf-3*(*pk1426*) (triangles) strains were infected with VSV-GFP, and GFP expression over time were analysed by confocal microscopy. **a**, Percentages of GFP-positive cells. **b**, Average intensities of GFP fluorescence per GFP-positive cell. Results are means $\pm$ s.e.m. for two to four independent fields at each time after infection.

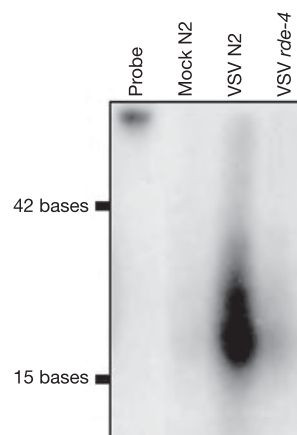


Figure 4 | Virus-specific siRNAs in infected N2 cells. RNAs from mock-infected N2, VSV-infected N2 and VSV-infected *rde-4* cells at 24 h after infection were analysed by RNase protection assays with the use of a radiolabelled probe specific for VSV N.

the host. The demonstration that VSV can infect *C. elegans* provides a new model system for identifying genes involved in regulating resistance or susceptibility not only to VSV but also to a wide range of viruses, using VSV-based vectors and pseudotyped viruses. The RNAi pathway is an example of such a set of host susceptibility determinants.

METHODS

Primary cell culture of *C. elegans*. Cells were isolated as described in ref. 27, with slight modifications. Early embryos were isolated after bleach treatment of cultures of synchronized adult hermaphrodites on sucrose step gradients. The embryos were incubated with chitinase (5 U ml⁻¹) and chymotrypsin (10 mg ml⁻¹). Monolayer cultures of the resultant dissociated cells were grown at 25 °C in a modified L15 medium containing 10% fetal bovine serum, 30 mM sucrose and 100 U ml⁻¹ penicillin/streptomycin.

VSV infection. Virus stocks were grown in baby hamster kidney (BHK-21) cells at low multiplicity (multiplicity of infection (m.o.i.) 0.01 plaque-forming units per cell) and infectious titres were measured by plaque assay on HeLa cells at 37 °C. *C. elegans* cultures were infected at high multiplicity (m.o.i. 10) by adding virus to the culture medium for 2 h at 25 °C and the virus inoculum was replaced with fresh medium. For virus growth curves, aliquots of the culture medium were taken at intervals of 6 h and viral titres were determined by plaque assay on BHK-21 cells.

Immunoprecipitation. Cell monolayers (4 × 10⁶ cells) were labelled from 42 to 48 h after infection with [³⁵S]methionine (100 µCi ml⁻¹), then harvested and lysed in RIPA buffer. Samples were incubated with monoclonal antibodies recognizing VSV N or G proteins, and Staph-G-conjugated beads were used to collect the antibody complexes. Samples were analysed by 10% SDS-PAGE and autoradiography.

RT-PCR analyses. RNA was isolated from infected cells with the use of Trizol. Equivalent amounts of total RNA were used in strand-specific RT-PCR reactions with specific RT primers that hybridized to viral sequences of either positive sense (N gene coding) or negative sense (P-M intergenic region). For the N gene, the forward primer is complementary to nucleotides 730–747 and the reverse primer is identical to nucleotides 1331–1365 of the minus-sense viral genome. For the P-M intergenic region, the forward primer is complementary to nucleotides 1499–1526 and the reverse primer was identical to nucleotides 2548–2570 of the viral genome. For the actin gene of *C. elegans*, the primer sequences are identical to nucleotides 124–146 (forward primer) and complementary to nucleotides 820–849 (reverse primer) of the coding sequence.

RNAse protection. Radiolabelled probes were generated by T7 transcription of cDNA plasmids of VSV N, P, M, G and L genes²⁸. RNA fractions enriched for small RNAs (less than 200 bases) were isolated from cells by using the mirVana miRNA isolation kit (Ambion) and RNAse protection assays used the mirVana miRNA detection kit (Ambion) in accordance with the manufacturer's protocols. Samples were analysed on a 15% denaturing polyacrylamide gel.

Confocal analysis. Cell monolayers were either examined for GFP fluorescence or processed for immunofluorescent staining by fixing them in 1% paraformaldehyde, permeabilizing them with 0.1% saponin plus 1% BSA and incubating them with primary antibodies against VSV-N or VSV-G proteins and phycoerythrin-conjugated secondary antibodies at appropriate dilutions.

Image analyses were performed with MetaMorph software. GFP images were thresholded and the percentage of GFP-positive cells and the intensity of GFP fluorescence in individual cells were measured.

Received 25 January; accepted 27 June 2005.

- Grishok, A. & Mello, C. C. RNAi (nematodes: *Caenorhabditis elegans*). *Adv. Genet.* **46**, 339–360 (2002).
- Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
- Tijsterman, M. & Plasterk, R. H. Dicars at RISC; the mechanism of RNAi. *Cell* **117**, 1–3 (2004).
- Denli, A. M. & Hannon, G. J. RNAi: an ever-growing puzzle. *Trends Biochem. Sci.* **28**, 196–201 (2003).
- Letchworth, G. J., Rodriguez, L. L. & Del cbarra, J. Vesicular stomatitis. *Vet. J.* **157**, 239–260 (1999).

- Rose, J. K. & Whitt, M. A. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 1221–1244 (Lippincott Williams & Wilkins, Philadelphia, 2004).
- Schnell, M. J., Buonocore, L., Kretzschmar, E., Johnson, E. & Rose, J. K. Foreign glycoproteins expressed from recombinant vesicular stomatitis viruses are incorporated efficiently into virus particles. *Proc. Natl Acad. Sci. USA* **93**, 11359–11365 (1996).
- Roberts, A. *et al.* Vaccination with a recombinant vesicular stomatitis virus expressing an influenza virus hemagglutinin provides complete protection from influenza virus challenge. *J. Virol.* **72**, 4704–4711 (1998).
- Boritz, E., Gerlach, J., Johnson, J. E. & Rose, J. K. Replication-competent rhabdoviruses with human immunodeficiency virus type 1 coats and green fluorescent protein: entry by a pH-independent pathway. *J. Virol.* **73**, 6937–6945 (1999).
- Takada, A. *et al.* A system for functional analysis of Ebola virus glycoprotein. *Proc. Natl Acad. Sci. USA* **94**, 14764–14769 (1997).
- Tabara, H. *et al.* The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
- Tijsterman, M., May, R. C., Simmer, F., Okihara, K. L. & Plasterk, R. H. Genes required for systemic RNA interference in *Caenorhabditis elegans*. *Curr. Biol.* **14**, 111–116 (2004).
- Tabara, H., Yigit, E., Siomi, H. & Mello, C. C. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DEXH-box helicase to direct RNAi in *C. elegans*. *Cell* **109**, 861–871 (2002).
- Simmer, F. *et al.* Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr. Biol.* **12**, 1317–1319 (2002).
- Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**, 645–649 (2004).
- Gitlin, L., Karelsky, S. & Andino, R. Short interfering RNA confers intracellular antiviral immunity in human cells. *Nature* **418**, 430–434 (2002).
- Ge, Q. *et al.* RNA interference of influenza virus production by directly targeting mRNA for degradation and indirectly inhibiting all viral RNA transcription. *Proc. Natl Acad. Sci. USA* **100**, 2718–2723 (2003).
- Hu, W. Y., Myers, C. P., Kilzer, J. M., Pfaff, S. L. & Bushman, F. D. Inhibition of retroviral pathogenesis by RNA interference. *Curr. Biol.* **12**, 1301–1311 (2002).
- Li, W. X. & Ding, S. W. Viral suppressors of RNA silencing. *Curr. Opin. Biotechnol.* **12**, 150–154 (2001).
- Vance, V. & Vaucheret, H. RNA silencing in plants—defense and counterdefense. *Science* **292**, 2277–2280 (2001).
- Voinnet, O. RNA silencing as a plant immune system against viruses. *Trends Genet.* **17**, 449–459 (2001).
- Li, H., Li, W. X. & Ding, S. W. Induction and suppression of RNA silencing by an animal virus. *Science* **296**, 1319–1321 (2002).
- Adelman, Z. N. *et al.* RNA silencing of dengue virus type 2 replication in transformed C6/36 mosquito cells transcribing an inverted-repeat RNA derived from the virus genome. *J. Virol.* **76**, 12925–12933 (2002).
- Bennasser, Y., Le, S. Y., Benkirane, M. & Jeang, K. T. Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. *Immunity* **22**, 607–619 (2005).
- Roth, B. M., Pruss, G. J. & Vance, V. B. Plant viral suppressors of RNA silencing. *Virus Res.* **102**, 97–108 (2004).
- Li, W. X. *et al.* Interferon antagonist proteins of influenza and vaccinia viruses are suppressors of RNA silencing. *Proc. Natl Acad. Sci. USA* **101**, 1350–1355 (2004).
- Christensen, M. *et al.* A primary culture system for functional analysis of *C. elegans* neurons and muscle cells. *Neuron* **33**, 503–514 (2002).
- Stillman, E. A., Rose, J. K. & Whitt, M. A. Replication and amplification of novel vesicular stomatitis virus minigenomes encoding viral structural proteins. *J. Virol.* **69**, 2946–2953 (1995).

Acknowledgements We thank the *Caenorhabditis* Genetics Center for most of the strains used in this study; M. Kaufmann for suggestions; and K. Mitchell for technical assistance in the initial phases of this study. This work was funded in part by UAMS Foundation research funds (M.C.), from the BRIN Program of the National Center for Research Resources (S.C.M. and M.C.), and startup funds (K.M.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to M.C. (chowmarie@uams.edu) or K.M. (kamachaca@uams.edu).

LETTERS

Conditional telomerase induction causes proliferation of hair follicle stem cells

Kavita Y. Sarin^{1,3}, Peggie Cheung¹, Daniel Gilson^{1,3}, Eunice Lee¹, Ruth I. Tennen^{1,2}, Estee Wang¹, Maja K. Artandi¹, Anthony E. Oro^{2,4} & Steven E. Artandi^{1,2}

TERT, the protein component of telomerase^{1,2}, serves to maintain telomere function through the *de novo* addition of telomere repeats to chromosome ends, and is reactivated in 90% of human cancers. In normal tissues, TERT is expressed in stem cells and in progenitor cells³, but its role in these compartments is not fully understood. Here we show that conditional transgenic induction of TERT in mouse skin epithelium causes a rapid transition from telogen (the resting phase of the hair follicle cycle) to anagen (the active phase), thereby facilitating robust hair growth. TERT overexpression promotes this developmental transition by causing proliferation of quiescent, multipotent stem cells in the hair follicle bulge region. This new function for TERT does not require the telomerase RNA component, which encodes the template for telomere addition, and therefore operates through a mechanism independent of its activity in synthesizing telomere repeats. These data indicate that, in addition to its established role in extending telomeres, TERT can promote proliferation of resting stem cells through a non-canonical pathway.

In stem cell and progenitor cell compartments^{3–5}, TERT has an important role in keeping telomeres sufficiently long and stable to prevent the adverse consequences of dysfunctional telomeres on cell viability and chromosomal stability^{6–8}. However, the need for expression of TERT in tissue stem cells and progenitor cells with long telomeres is less clear, especially in laboratory mice, whose telomeres are significantly longer than those of humans (40–60 kilobases (kb) versus 5–15 kb). Moreover, recent findings indicate that TERT promotes tumour development even in settings of ample telomere reserve, although the mechanisms underlying these telomere length-independent activities of TERT remain unclear^{9–13}. We therefore hypothesized that TERT may exert effects in stem cell and progenitor cell compartments that could explain both its regulation during lineage development and its poorly understood telomere length-independent activities.

To test this hypothesis, we turned to the mammalian hair follicle, an organ that harbours tightly regulated multipotent stem cells and cycles between telogen and anagen¹⁴. Initiation of a new anagen cycle depends on activation of a small number of quiescent stem cells that reside in the bulge, a niche at the follicle base^{15–19}. These activated stem cells proliferate and differentiate into progenitor cells (matrix cells) that give rise to the differentiated lineages that comprise the hair shaft and root sheaths. This period of hair synthesis ceases when the new section of the anagen follicle is remodelled through apoptotic regression (catagen), resulting in another telogen phase (Fig. 1a). To understand how telomerase is regulated during mouse hair follicle cycling, we analysed telomerase activity in mouse skin, exploiting the fact that hair follicle cycles are synchronized for the first two

postnatal periods of hair follicle growth, approximately 8 weeks²⁰. Protein extracts from wild-type mouse skin were obtained between postnatal days 4 and 52, to allow analysis of both the first and second postnatal hair cycles using the telomere repeat amplification protocol (TRAP assay). Telomerase activity strongly correlated with the first and second anagen phases and was not detected during telogen phases (Fig. 1b). These data indicate that, in mouse hair follicles, as in human²¹, telomerase is associated with the anagen phase, a period of intense progenitor cell activity.

To determine whether TERT can modulate adult tissue stem cell and progenitor cell function, we engineered a conditional TERT transgenic system in mice using a tetracycline-regulated approach²². The mouse TERT complementary DNA was cloned under control of a tetracycline responsive promoter (tetop-TERT⁺). To drive expression of TERT, we expressed the reverse tetracycline transactivator (rtTA), which binds to and activates the tetracycline responsive promoter, using a cytomegalovirus (CMV) enhancer/ β -actin promoter (actin-rtTA⁺) because this element was previously shown to be active in stem cells²³ and in many epithelial tissues, including skin^{24,25}. Tetop-TERT⁺ mice were intercrossed with actin-rtTA⁺ mice to generate actin-rtTA⁺;tetop-TERT⁺ (termed inducible TERT or i-TERT) mice. To induce expression of TERT, i-TERT mice were administered drinking water containing the tetracycline analogue doxycycline. Northern blot and TRAP assay confirmed doxycycline-dependent induction of the TERT transgene and increased telomerase activity in the skin of i-TERT mice (Fig. 1c, d) as well as in other tissues (data not shown). Remarkably, within several weeks of doxycycline treatment, the coats of i-TERT mice were markedly altered, reminiscent of mice with known mutations that affect hair follicle cycling^{26,27} (Fig. 1e).

To determine whether abnormalities in hair follicle cycling might underlie the altered hair phenotype, we analysed skin biopsies from i-TERT mice after doxycycline administration beginning at day 21. Hair follicles were appropriately in anagen at day 28 in i-TERT mice on and off doxycycline treatment, and in littermate controls. By day 50, follicles from i-TERT mice off doxycycline treatment and from non-transgenic mice had exited anagen and were in the second postnatal telogen phase. In marked contrast, hair follicles from i-TERT mice on doxycycline were consistently in anagen at day 50 (Fig. 1f). This effect was doxycycline-dependent and occurred with 100% penetrance in i-TERT mice (18 out of 18 in anagen) (χ^2 analysis, $P < 0.0001$ for i-TERT on versus off doxycycline; see also Supplementary Table S1). RNA *in situ* hybridization revealed a pan-epithelial expression pattern of transgenic TERT in skin that included the keratin-14⁺ compartment, but spared the dermal papilla, indicating that transgenic TERT messenger RNA is expressed principally in hair follicle and skin epithelium (Fig. 1g, h). Together, these data

¹Department of Medicine, Division of Hematology, ²Cancer Biology Program, ³Department of Genetics and ⁴Department of Dermatology, Stanford School of Medicine, Stanford, California 94305, USA.

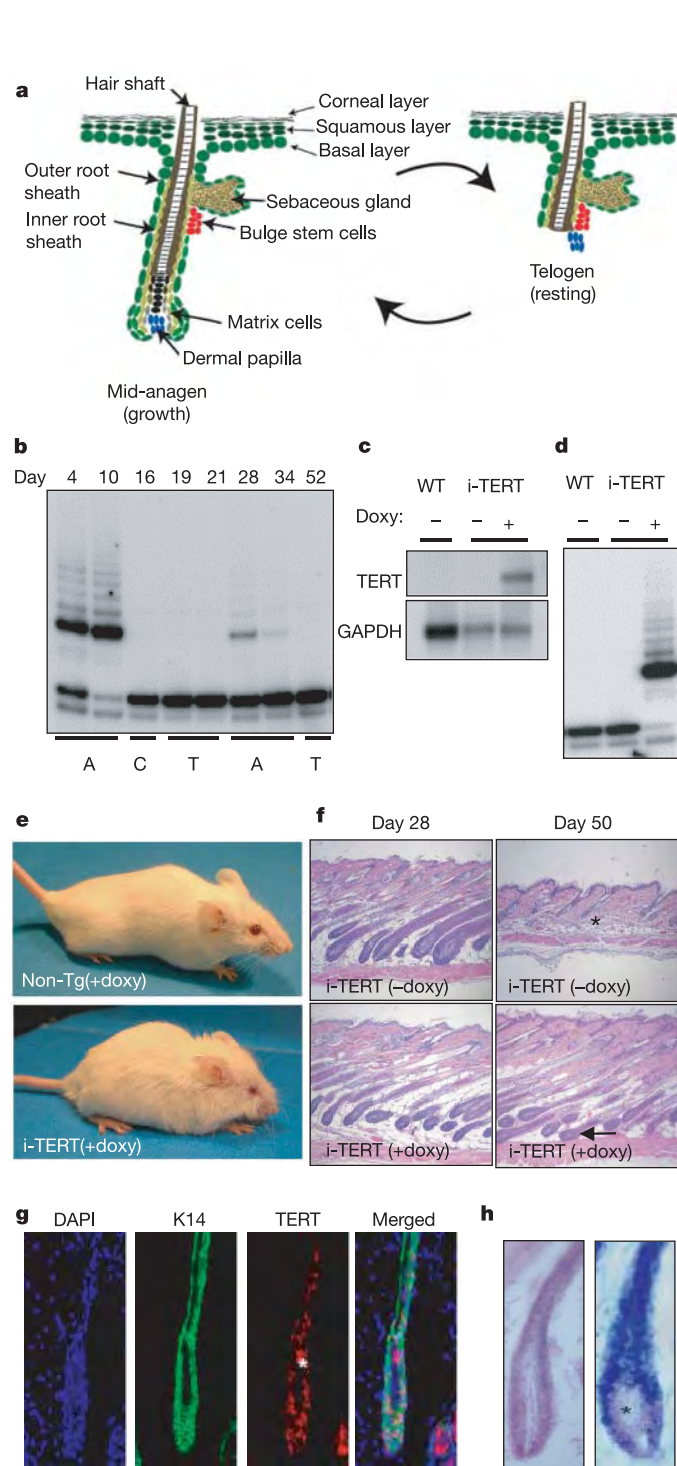


Figure 1 | Conditional activation of TERT promotes the anagen phase of the hair follicle cycle. **a**, A schematic diagram of hair follicle cycling. **b**, TRAP on non-transgenic (Non-Tg) skin samples at day 4, 10, 16, 19, 21, 28, 34 and 52 (anagen, A; catagen, C; telogen, T). **c**, **d**, Northern blot analysis and TRAP on skin extracts from i-TERT and wild type (WT) mice at day 50. **e**, Doxycycline-treated mice at day 70, Non-Tg (+doxy) and i-TERT (+doxy) mice. **f**, H&E skin sections at $\times 20$ magnification. -doxy indicates no doxycycline treatment; asterisk indicates telogen hair follicle; arrow indicates anagen hair follicle. **g**, RNA *in situ* hybridization for TERT mRNA and immunofluorescence for K14 in i-TERT (+doxy) skin (asterisk indicates autofluorescence). **h**, RNA *in situ* hybridization for TERT mRNA (blue) in an i-TERT(+doxy) skin section (right panel) and WT anagen skin section (left panel) (asterisk indicates dermal papilla).

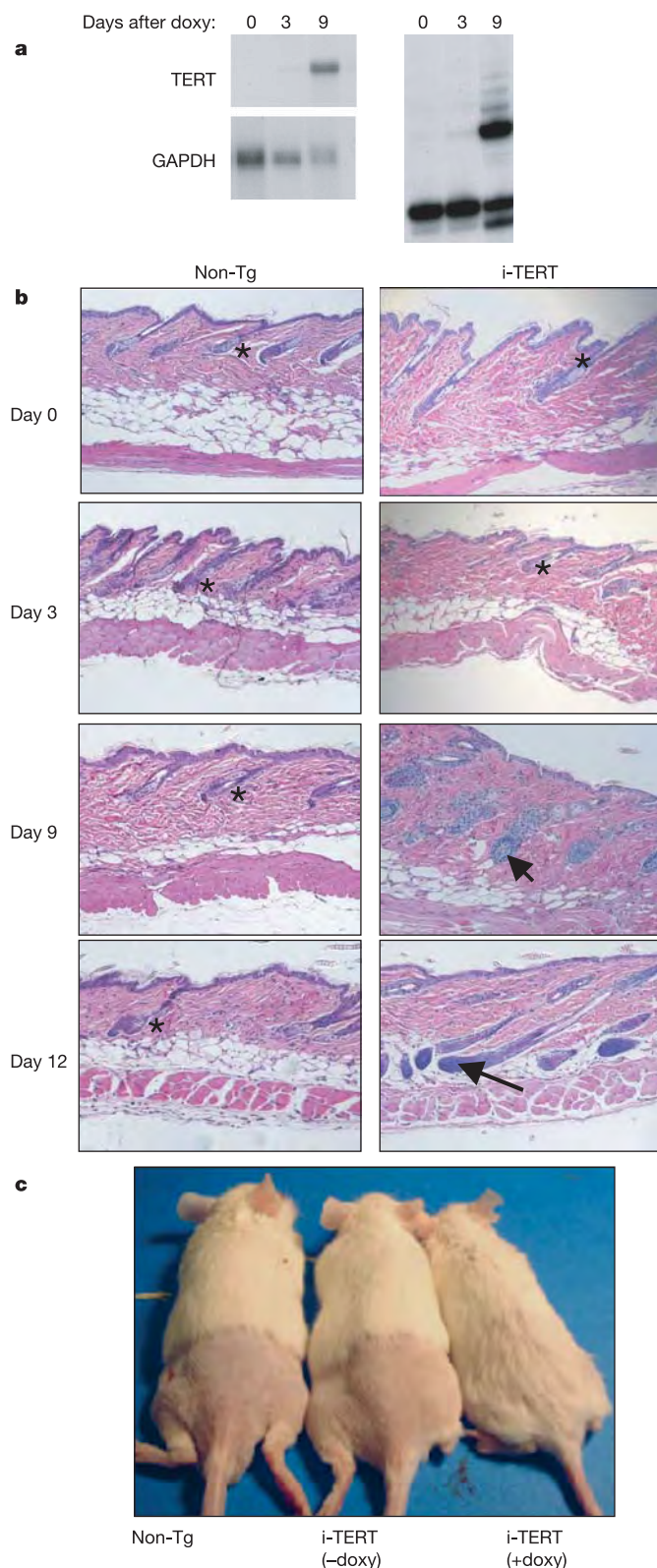


Figure 2 | TERT induction triggers a rapid transition from telogen to anagen. **a**, Northern blot (left) and TRAP assay (right) show increased TERT expression and telomerase activity after nine days of doxycycline treatment. **b**, H&E stain showing that follicles in i-TERT mice entered anagen (black arrows) by day 9, whereas Non-Tg controls remained in telogen (asterisks). **c**, Hair growth was observed only in i-TERT mice with doxycycline treatment (+doxy), but not in i-TERT mice (-doxy) or Non-Tg littermates.

show that conditional induction of TERT in adult hair follicle epithelium promotes the anagen phase.

To determine whether expression of TERT is sufficient to induce a transition from telogen to anagen, i-TERT mice were treated with doxycycline after hair follicles had entered the prolonged second telogen (day 40 in FVB mice; Fig. 2), and biopsies were taken at regular intervals for 12 days. TERT mRNA and telomerase activity progressively increased from day 3 to day 9 (Fig. 2a). Whereas hair follicles from non-transgenic mice remained in telogen for the duration of the time course, follicles from i-TERT mice treated with doxycycline entered anagen by day 9 and were in mid-anagen²⁰ by day 12 (Fig. 2b) (χ^2 analysis, $P = 0.005$; see also Supplementary Table S2). The kinetics of anagen entry closely followed the timing of TERT induction. To determine whether induction of anagen by TERT enabled hair growth, i-TERT mice were treated with doxycycline containing drinking water beginning at day 45, and shaved at day 55. Fourteen days after shaving, i-TERT mice administered doxycycline showed significant hair growth as compared with both i-TERT mice off doxycycline and non-transgenic controls in which hair did not grow during this interval (Fig. 2c). Notably, these results show that induction of TERT in telogen follicles is sufficient to initiate a transition to the anagen phase and promote new hair synthesis.

Because activation of bulge stem cells is integral to the initiation of a new anagen cycle^{15,16,18}, we hypothesized that TERT's effects on the hair follicle cycle might be mediated through the stem cell compartment. To address this hypothesis, we used a label retaining technique that marks hair follicle bulge stem cells by repeated injections of BrdU followed by a long chase period¹⁵. Cohorts of i-TERT mice and non-transgenic controls were injected with BrdU at 10 days of age (day 10). During the second telogen, mice in each group were

biopsied, switched to doxycycline containing drinking water and biopsied again between days 80 and 100. Label retaining cells (LRCs) were visualized by double immunostaining with antibodies against BrdU and CD34, a cell membrane marker for hair follicle stem cells^{28,29}. LRCs were present in similar numbers in both i-TERT and non-transgenic mice at day 55, before the switch to doxycycline containing water (approximately 0.6 BrdU⁺ cell per CD34⁺ cell). After five weeks of doxycycline treatment, BrdU label in CD34⁺ stem cells was retained in non-transgenic mice at comparable levels, consistent with previous observations that BrdU label persists in quiescent bulge cells for more than 6 months¹⁷. In marked contrast, BrdU label was profoundly depleted in the CD34⁺ cell population in the bulge by induction of TERT in i-TERT mice (Fig. 3a, b; 76% reduction in BrdU⁺ cell per CD34⁺ cell; $P < 0.0001$). Despite the loss of BrdU label, CD34⁺ cells in the bulge remained in similar numbers, indicating that, under the influence of TERT, stem cells divide but probably self-renew to maintain the CD34⁺ population. A similar reduction in LRCs in i-TERT mice was seen in epidermal wholemounts, corroborating the effects seen in dorsal skin sections (Fig. 3c). These data show that TERT causes hair follicle bulge cells to proliferate, diluting BrdU label from this quiescent stem cell population.

To determine whether TERT broadly enhances keratinocyte proliferation, we measured the proliferation index in the basal layer of the interfollicular epidermis (Fig. 3d). Despite expression of transgenic TERT mRNA in this compartment, proliferation was not substantially altered in the basal layer in i-TERT mice compared to non-transgenic littermates in anagen (Fig. 3e; 4.2 Ki-67⁺ cells per 100 μ m length of epidermis for i-TERT day 50 compared to 4.3 Ki-67⁺ cells per 100 μ m for non-transgenic day 28). Furthermore, we found no changes in structure, differentiation or signalling in either

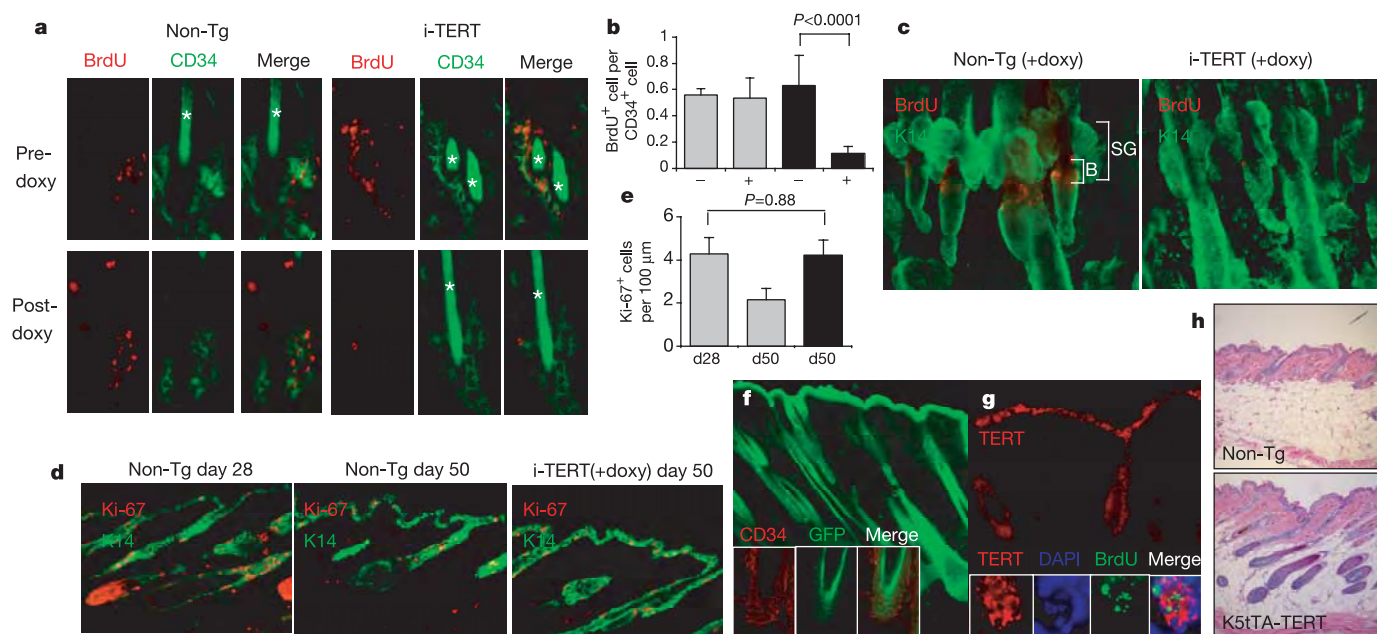


Figure 3 | TERT activates stem cells, depleting BrdU label from LRCs.

a, Immunofluorescence for BrdU (red) and CD34 (green) shows maintenance of LRCs in non-transgenic (Non-Tg) group, but dramatic loss of label in i-TERT mice after doxycycline (doxy) treatment (pre-doxycy, day 55; post-doxycy, day 90). Asterisk indicates autofluorescence of hair. **b**, Quantification of LRC data from **a**, showing the fraction of CD34⁺ cells that are also BrdU⁺. Data for i-TERT mice (black bars, $n = 4$ mice) and non-transgenic mice (grey bars, $n = 3$ mice), pre-doxycy (–) and post-doxycy (+) treatment are shown. **c**, LRC analysis from wholemounts of epidermis from the tail of mice labelled with BrdU at day 10, switched to doxy at day 40 and analysed at day 65 (BrdU, red; K14, green); B indicates bulge and SG

indicates sebaceous gland. **d**, Immunofluorescence using Ki-67 (red) to mark proliferating cells and K14 (green) to identify basal layer of skin. **e**, Quantification of proliferation index in **d** as Ki-67⁺ cells per 100 μ m length of basal layer (each comparison $n = 2$ mice). i-TERT mice (black bar) and non-transgenic mice (grey bars). **f**, GFP epifluorescence co-stained with CD34 (inset, confocal microscopy) in skin section from an actin-GFP mouse. **g**, RNA *in situ* analysis for TERT mRNA in i-TERT(+doxy) mouse skin (inset, TERT mRNA expression (cytoplasmic) overlaps with LRCs, marked by BrdU (nuclear)). **h**, H&E sections from K5tTA-TERT⁺(–doxy) (bottom) and Non-Tg (top) mice, at $\times 20$ magnification. Error bars indicate standard deviation. P values derived from Student's *t*-test.

hair follicle or interfollicular epidermis in i-TERT mice (see Supplementary Figs S2 and S3). We therefore conclude that the principal effects of TERT in this system occur through activation of quiescent hair follicle stem cells.

To understand whether these results are consistent with a direct effect of TERT on stem cells, we asked whether transgenic TERT is expressed in the stem cell compartment. We found that the actin promoter element used to direct rtTA expression is strongly active in CD34⁺ bulge cells in actin–green fluorescent protein (GFP) transgenic mice²⁴ (Fig. 3f). Furthermore, TERT mRNA was co-expressed with BrdU in LRCs in the bulge region in i-TERT mice (Fig. 3g). The induction of anagen can occur through signals from the dermal papilla³⁰, but the lack of detectable levels of TERT mRNA in the dermal papilla (Fig. 1h) makes it unlikely in this case. To confirm that TERT exerts its effect through the epithelium, we intercrossed tetop-TERT mice with a transgenic mouse in which the keratin-5 promoter drives expression of the tetracycline transactivator (tTA) in the basal layer and outer root sheath (K5-tTA, activity inhibited by tetracyclines)¹⁸. Compound K5-tTA⁺;tetop-TERT⁺ mice were bred on doxycycline and weaned off doxycycline-containing drinking water

at day 21 to induce the TERT transgene. Expression of TERT mRNA in skin epithelium (data not shown) induced anagen in 5 out of 5 K5-tTA⁺;tetop-TERT⁺ mice, whereas all littermate control biopsies were in telogen (Fig. 3h; 6 out of 6; χ^2 analysis $P = 0.0009$). These data show that TERT's effects in promoting anagen are intrinsic to the keratin-5 compartment of the skin epithelium, the layer where the hair follicle stem cells reside.

These effects of TERT in facilitating a switch from telogen to anagen might occur either through TERT's established role in telomere synthesis or through a new mechanism, independent of its known enzymatic function. Telomere synthesis requires both TERT and TERC (telomerase RNA component); therefore, if the effects of TERT are retained in a TERC^{-/-} background, telomere extension cannot be required for stem cell activation events mediated by TERT. To answer this question, TERC^{+/-} mice were intercrossed with inducible TERT alleles to derive cohorts of i-TERT mice that were TERC^{+/+}, TERC^{+/-} and TERC^{-/-}. Mice in each group were treated with doxycycline beginning in telogen at day 40. Histological analysis at day 55 revealed that conditional activation of TERT induced anagen in 5 out of 5 i-TERT TERC^{-/-} mice (Fig. 4a; $P = 0.0003$ for i-TERT TERC^{-/-} mice on doxycycline versus i-TERT TERC^{-/-} mice off doxycycline; see also Supplementary Table S3). These results were identical to those obtained from i-TERT mice in either TERC^{+/+} or TERC^{+/-} backgrounds (6 out of 6 in anagen), indicating that TERC is not required for TERT to induce anagen. TRAP assays and RNA analysis by reverse transcription–polymerase chain reaction (RT–PCR) confirmed the absence of telomerase activity and TERC RNA in the skin of i-TERT TERC^{-/-} mice (Fig. 4b). These data prove that the mechanism by which TERT triggers hair follicles to transition from telogen to anagen does not require telomere synthesis.

Here we show that conditional activation of TERT protein in mouse skin is sufficient to induce proliferation of hair follicle stem cells, and that this function is independent of its role in telomere synthesis. In this transgenic context, either altered timing of TERT expression or higher protein levels may be important in mediating these new activities for TERT. Our data encourage further studies to understand the role of the endogenous TERT protein in stem cells, including loss-of-function studies and experiments designed to determine whether TERT's function in stem cell proliferation is cell autonomous. We propose that as TERT is dynamically regulated during the course of stem and progenitor cell maturation, or during advancing stages of cancer development, TERT may directly support these processes. These data provide a novel framework for understanding the recently identified telomere length-independent functions of TERT^{9–12}, and suggest new strategies for manipulating TERT for therapeutic purposes in treating disorders associated with tissue injury and ageing.

METHODS

Generation of transgenic mice. TERT was placed under control of a tetracycline-inducible promoter by subcloning a 3.5 kb *EcoRI* fragment of the mouse TERT cDNA into the *EcoRI* site of pUHD10-3 (ref. 22). To create actin-rtTA, an *EcoRI*–*BamHI* fragment of the rtTA cDNA²² was subcloned into the *EcoRI* site of pCAGS²⁴ by blunt-ended ligation. Prokaryotic sequences were excised from each plasmid and the gel-isolated DNA fragments were separately injected into pronuclei of FVB/N fertilized zygotes. Founder mice were screened by PCR and Southern blot analysis. Actin-rtTA transgene-positive mice were intercrossed with tetop-TERT transgene-positive mice to generate i-TERT double transgenic mice for characterization. Doxycycline (2 mg ml⁻¹ in 5% sucrose for rtTA, 10 μ g ml⁻¹ in water for tTA) was administered in drinking water in light-protected bottles and changed biweekly. TERC^{+/-} mice were backcrossed to the FVB/N strain for six generations, then intercrossed with the i-TERT alleles to yield i-TERT mice on TERC^{+/+}, TERC^{+/-} and TERC^{-/-} backgrounds. All mice were treated in accordance with AAALAC approved guidelines at Stanford University.

Histology. Skin biopsies were obtained from dorsal skin of mice under anaesthesia and wounds were closed with 9-mm staples (Roboz). Biopsy specimens were either fixed overnight in 10% formalin then embedded in paraffin or

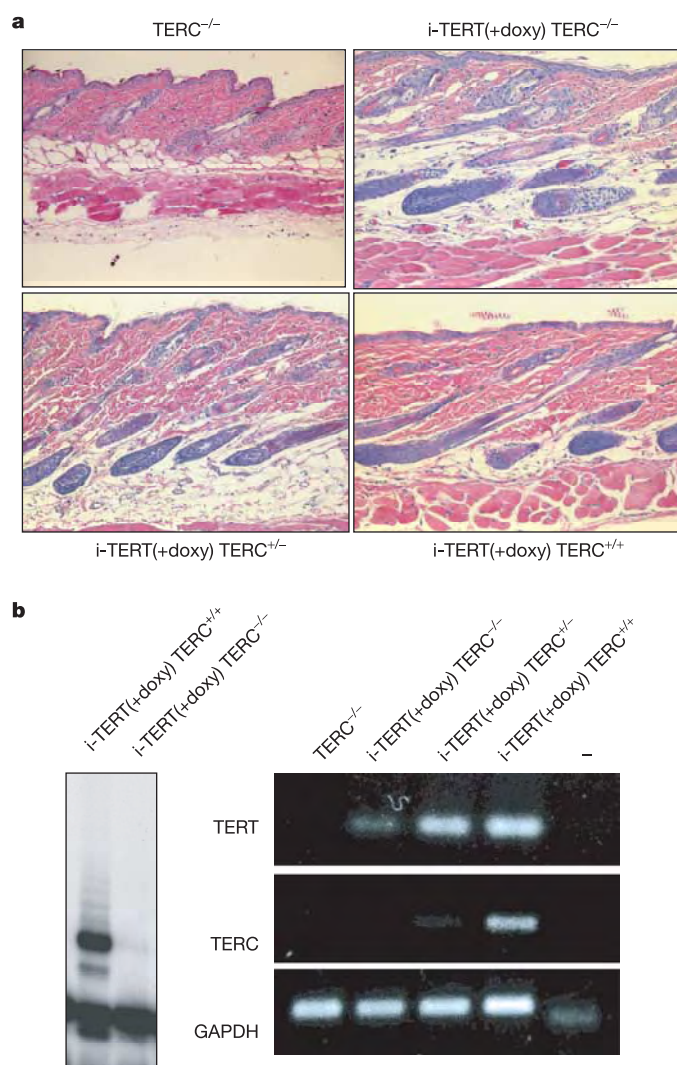


Figure 4 | TERT's activity in facilitating a transition from telogen to anagen is independent of its function in telomere synthesis. **a**, H&E sections showing TERT-induced anagen in mice with TERC^{+/+}, TERC^{+/-} and TERC^{-/-} backgrounds, at $\times 20$ magnification. **b**, Skin samples from i-TERT TERC^{-/-} mice (+doxy) lacked telomerase activity by TRAP (left panel) and lacked TERC expression by RT–PCR (right panel). Minus sign indicates negative control lacking reverse transcriptase.

directly embedded in OCT and frozen on an isopropanol-dry ice slurry. Formalin-fixed, paraffin-embedded tissue sections were stained with haematoxylin and eosin (H&E) for microscopic analysis.

RNA *in situ* analysis. Digoxigenin-labelled anti-sense RNA probes were synthesized *in vitro* using digoxigenin-UTP (Roche). *In situ* analysis was performed on 10 µm frozen sections or 5 µm paraffin sections. Probes were detected using a horseradish peroxidase conjugated anti-digoxigenin antibody and the slides were developed using biotinyl tyramide (Dako) followed by Cy3-Streptavidin (Vector Laboratories). Slides were mounted with Vectashield plus DAPI (Vector Laboratories). For chromogenic RNA *in situ* analysis, probes were detected using an alkaline phosphatase conjugated anti-digoxigenin antibody and were developed using NBT/BCIP (Roche) and counterstained with nuclear fast red (Vector Laboratories).

Immunofluorescence. All assays were performed on 10 µm frozen sections. Samples were fixed in 10% formalin then blocked with either 10% normal goat serum in TBS-T or MOM IgG Blocking solution (Vector laboratories). Sections were incubated in primary antibody overnight at 4 °C. Primary antibodies included: mouse anti-Ki-67 (Pharmingen), rabbit anti-K14 (Covance), rat anti-BrdU (BD) and rat-anti-CD34 (Pharmingen). For BrdU detection, slides were pre-treated in 1 M HCl for 1 h at 37 °C.

Analysis of label retaining cells. To label follicle stem cells, 10-day-old mice were injected with 250 µg of BrdU every 12 h for four injections to mark proliferating epidermal keratinocytes. Skin samples were obtained from the mice after an extended chase period of 45–90 days. BrdU immunofluorescence was performed on frozen sections to visualize label retaining cells, followed by co-staining for CD34.

Tail wholemount immunolabelling. Wholemounts of tail epidermis were prepared and stained for BrdU and K14 as described¹⁷.

Northern blots, RT-PCR and telomerase activity assays. Tissues were snap frozen in liquid nitrogen and then ground with mortar and pestle. RNA was isolated from organs or cells by means of homogenization in Trizol. 5 µg of total RNA was fractionated on a 0.8% formaldehyde gel, transferred to Hybond-N membrane (Amersham) and hybridized with TERT or GAPDH ³²P-labelled DNA probes. For RT-PCR, the reverse transcriptase reaction was performed on 1 µg of total RNA using Superscript II, and PCR was performed with primer pairs specific for TERT or GAPDH. For TRAP assays, protein was extracted from 50–100 mg of tissue in CHAPS lysis buffer and a standard TRAP reaction was performed (TRAPEZE, Chemicon).

Received 13 February; accepted 6 May 2005.

- Blackburn, E. H. Switching and signalling at the telomere. *Cell* **106**, 661–673 (2001).
- Smogorzewska, A. & De Lange, T. Regulation of telomerase by telomeric proteins. *Annu. Rev. Biochem.* **73**, 177–208 (2004).
- Morrison, S. J., Prowse, K. R., Ho, P. & Weissman, I. L. Telomerase activity in hematopoietic cells is associated with self-renewal potential. *Immunity* **5**, 207–216 (1996).
- Yui, J., Chiu, C. P. & Lansdorf, P. M. Telomerase activity in candidate stem cells from fetal liver and adult bone marrow. *Blood* **91**, 3255–3262 (1998).
- Forsyth, N. R., Wright, W. E. & Shay, J. W. Telomerase and differentiation in multicellular organisms: turn it off, turn it on, and turn it off again. *Differentiation* **69**, 188–197 (2002).
- Blasco, M. A. *et al.* Telomere shortening and tumour formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
- Lee, H. W. *et al.* Essential role of mouse telomerase in highly proliferative organs. *Nature* **392**, 569–574 (1998).
- Rudolph, K. L. *et al.* Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* **96**, 701–712 (1999).
- Gonzalez-Suarez, E. *et al.* Increased epidermal tumors and increased skin wound healing in transgenic mice overexpressing the catalytic subunit of telomerase, mTERT, in basal keratinocytes. *EMBO J.* **20**, 2619–2630 (2001).
- Artandi, S. E. *et al.* Constitutive telomerase expression promotes mammary carcinomas in aging mice. *Proc. Natl Acad. Sci. USA* **99**, 8191–8196 (2002).
- Stewart, S. A. *et al.* Telomerase contributes to tumorigenesis by a telomere length-independent mechanism. *Proc. Natl Acad. Sci. USA* **99**, 12606–12611 (2002).
- Chang, S., Khoo, C. M., Naylor, M. L., Maser, R. S. & DePinho, R. A. Telomere-based crisis: functional differences between telomerase activation and ALT in tumour progression. *Genes Dev.* **17**, 88–100 (2003).
- Blasco, M. A. & Hahn, W. C. Evolving views of telomerase and cancer. *Trends Cell Biol.* **13**, 289–294 (2003).
- Alonso, L. & Fuchs, E. Stem cells in the skin: waste not, Wnt not. *Genes Dev.* **17**, 1189–1200 (2003).
- Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61**, 1329–1337 (1990).
- Taylor, G., Lehrer, M. S., Jensen, P. J., Sun, T. T. & Lavker, R. M. Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell* **102**, 451–461 (2000).
- Braun, K. M. *et al.* Manipulation of stem cell proliferation and lineage commitment: visualisation of label-retaining cells in wholemounts of mouse epidermis. *Development* **130**, 5241–5255 (2003).
- Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).
- Morris, R. J. *et al.* Capturing and profiling adult hair follicle stem cells. *Nature Biotechnol.* **22**, 411–417 (2004).
- Muller-Rover, S. *et al.* A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. *J. Invest. Dermatol.* **117**, 3–15 (2001).
- Ramirez, R. D., Wright, W. E., Shay, J. W. & Taylor, R. S. Telomerase activity concentrates in the mitotically active segments of human hair follicles. *J. Invest. Dermatol.* **108**, 113–117 (1997).
- Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl Acad. Sci. USA* **89**, 5547–5551 (1992).
- Wright, D. E. *et al.* Cyclophosphamide/granulocyte colony-stimulating factor causes selective mobilization of bone marrow hematopoietic stem cells into the blood after M phase of the cell cycle. *Blood* **97**, 2278–2285 (2001).
- Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).
- Sawamura, D. *et al.* Promoter/enhancer cassettes for keratinocyte gene therapy. *J. Invest. Dermatol.* **112**, 828–830 (1999).
- Hebert, J. M., Rosenquist, T., Gotz, J. & Martin, G. R. GF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations. *Cell* **78**, 1017–1025 (1994).
- Gat, U., DasGupta, R., Degenstein, L. & Fuchs, E. De Novo hair follicle morphogenesis and hair tumors in mice expressing a truncated beta-catenin in skin. *Cell* **95**, 605–614 (1998).
- Trempey, C. S. *et al.* Enrichment for living murine keratinocytes from the hair follicle bulge with the cell surface marker CD34. *J. Invest. Dermatol.* **120**, 501–511 (2003).
- Blanpain, C., Lowry, W. E., Geoghegan, A., Polak, L. & Fuchs, E. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell* **118**, 635–648 (2004).
- Sato, N., Leopold, P. L. & Crystal, R. G. Induction of the hair growth phase in postnatal mice by localized transient expression of Sonic hedgehog. *J. Clin. Invest.* **104**, 855–864 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We wish to acknowledge technical assistance from the Stanford Transgenic Core Facility and from P. Chu in the Stanford Comparative Medicine Histology Research Core Laboratory. We wish to thank R. DePinho for the use of TERC^{−/−} mice, A. Glick for the use of K5-tTA mice, T. Sun for the gift of AE13 and AE15 antibodies, and K. Braun and V. Horsley for technical insights. We appreciate comments and insights from T. de Lange, R. Nusse, I. Weissman, L. Attardi, J. Sage, A. Brunet, M. Cleary, D. Felsner, P. Khavari and members of the Artandi laboratory. K.Y.S. is supported by a Medical Scientist Training Program Grant. R.I.T. is supported by a Stanford Graduate Fellowship and a NSF Fellowship. A.E.O. is supported by NIAMS. This work was supported by grants from the Rita Allen Foundation, the V Foundation, the NIH and the Stanford Cancer Council to S.E.A.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.E.A. (sartandi@stanford.edu).

Endonucleolytic processing of covalent protein-linked DNA double-strand breaks

Matthew J. Neale¹, Jing Pan¹ & Scott Keeney^{1,2}

DNA double-strand breaks (DSBs) with protein covalently attached to 5' strand termini are formed by Spo11 to initiate meiotic recombination^{1,2}. The Spo11 protein must be removed for the DSB to be repaired, but the mechanism for removal is unclear³. Here we show that meiotic DSBs in budding yeast are processed by endonucleolytic cleavage that releases Spo11 attached to an oligonucleotide with a free 3'-OH. Two discrete Spo11-oligonucleotide complexes were found in equal amounts, differing with respect to the length of the bound DNA. We propose that these forms arise from different spacings of strand cleavages flanking the DSB, with every DSB processed asymmetrically. Thus, the ends of a single DSB may be biochemically distinct at or before the initial processing step—much earlier than previously thought. SPO11-oligonucleotide complexes were identified in extracts of mouse testis, indicating that this mechanism is evolutionarily conserved. Oligonucleotide-topoisomerase II complexes were also present in extracts of vegetative yeast, although not subject to the same genetic control as for generating Spo11-oligonucleotide complexes. Our findings suggest a general mechanism for repair of protein-linked DSBs.

We previously proposed that Spo11 might be removed from DSB ends by either of two mechanisms: direct hydrolysis of the covalent protein-DNA linkage, or single-stranded endonucleolytic cleavage releasing Spo11 covalently attached to a short oligonucleotide^{1,4} (Fig. 1a). These mechanisms are distinguished by the presence or absence of an oligonucleotide bound to Spo11. We identified this predicted protein-DNA complex using a direct biochemical approach in *Saccharomyces cerevisiae*. A strain expressing haemagglutinin (HA) epitope-tagged Spo11 was induced to enter meiosis, denaturing extracts were prepared and then Spo11-HA was immunoprecipitated and treated with ³²P-labelled nucleotide and terminal deoxynucleotidyl transferase (TdT), which catalyses untemplated addition of nucleotides to DNA with a free 3'-OH end. A chain terminating nucleotide was used to limit incorporation to a single residue.

Four radiolabelled bands were observed between ~60 and 110 kDa (Fig. 1b, lane 3). Two bands (marked with asterisks) were non-specific because they were present when TdT and nucleotide were incubated alone (Fig. 1b, lane 1). Two bands (indicated by filled arrowheads) were specific for Spo11-HA because they were not seen with mock immunoprecipitation (Fig. 1b, lane 2) or untagged Spo11 (Fig. 1b, lane 4). Labelling was not observed if DSBs were not formed, namely, when the catalytic tyrosine of Spo11 was mutated to phenylalanine² (Fig. 1b, lane 5) or in a *mei4* mutant⁴ (Fig. 1b, lane 6). Labelled Spo11 species were also not observed in *rad50S* or *sae2Δ* mutants (Fig. 1b, lanes 7 and 8), in which Spo11 remains covalently attached to DSB ends^{1,5}. Thus, Spo11-oligonucleotide complexes did not arise from non-physiological disruption of covalent Spo11-DSB complexes.

Oligonucleotide-associated Spo11 was compared to the free protein by combined autoradiography and western blotting analysis (Fig. 1c). The predominant signal from western blotting (Fig. 1c, panel 2) was from free Spo11-HA (~55 kDa, open arrow). Bands co-migrating with the radiolabelled species were not observed after a short exposure of the western blot (compare panels 1 and 2, Fig. 1c), indicating that the majority of Spo11 was not DNA-associated. Moreover, Spo11 that made DSBs in *rad50S* and *sae2Δ* mutants remains attached to high molecular weight DNA and is not resolved on SDS polyacrylamide gel electrophoresis (SDS-PAGE)¹, but free Spo11 levels were not detectably reduced in these mutants relative to strains with no DSBs (Fig. 1b, lower panel). Thus, only a minority of Spo11 protein engages in DSB formation⁶. Longer exposure of the western blot revealed slower migrating bands, one co-migrating with the upper radioactive signal (Fig. 1c, panel 3). Free Spo11-HA presumably masked the western blot signal from the lower radioactively labelled band. After partial fading of the chemiluminescent signal and re-exposure of the blot, free Spo11-HA provided the only signal visible by chemiluminescence, superimposed on the autoradiographic signal (Fig. 1c, panel 4). The larger and smaller labelled species were shifted by +15 and +5 kDa, respectively.

Two discrete labelled species were not expected. To determine their nature, they were gel-purified and digested with Pronase and the DNA fragments were separated on a sequencing gel (Fig. 1d). The DNA from the upper Spo11 band migrated from ~24–40 nucleotides, whereas DNA from the lower band ran from ~10–15 nucleotides. The smeared migration patterns probably arise from heterogeneities in oligonucleotide lengths, DNA sequence and the completeness of protease digestion. Estimating true oligonucleotide size requires correction for radiolabelled nucleotide incorporation and for any amino acids left after protease digestion⁷. Also, control experiments indicated that oligonucleotides <10 nucleotides were inefficiently recovered (data not shown), so the lower limit for the smaller oligonucleotides is not accurately known. We estimate that the two complexes contain oligonucleotides of ~21–37 nucleotides and ≤12 nucleotides, respectively.

If Spo11-oligonucleotide complexes are products of DSB processing, then they should appear with similar kinetics as resected (Spo11-free) DSBs. In a synchronously sporulating culture, Spo11-oligonucleotide complexes were detected as early as two hours after transfer to sporulation medium, accumulated to peak levels at four hours and declined thereafter (Fig. 1e, f). As predicted, the appearance of these complexes closely matched the timing for resected DSBs in the same culture (Fig. 1f). The long and short-oligonucleotide forms were consistently in a 1:1 ratio (long:short = 1.03 ± 0.10, mean ± s.d., *n* = 6 cultures). The ratio did not vary significantly with changes in amount of TdT, labelling time or dCTP versus cordycepin triphosphate as the source of label (data not shown),

¹Molecular Biology Programs, Memorial Sloan-Kettering Cancer Center and ²Weill Graduate School of Medical Sciences of Cornell University, 1275 York Avenue, New York, New York 10021, USA.

suggesting that this measurement is a faithful reporter of the relative amounts of the two species.

The Mre11–Rad50–Xrs2 complex is probably the nuclease generating Spo11–oligonucleotide complexes because Mre11 has single-strand endonuclease activity, and certain *rad50* and *mre11* mutants (including nuclease-defective alleles of *mre11*) cannot remove Spo11 from DSB ends (reviewed in ref. 4). Phosphatase treatment before incubation with TdT did not increase labelling (data not shown), indicating that the Spo11-associated oligonucleotides generated *in vivo* have free 3'-OH termini, consistent with Mre11 nuclease activity. We propose that Mre11 directly cleaves single strands adjacent to Spo11-associated DSBs (see below). However, it is also possible that Spo11 is released by 3'-5' exonuclease processing of more distantly placed single-strand breaks. Such a mechanism is consistent with the polarity of weak Mre11 exonuclease activity⁸, but would require that Mre11 exonuclease work much more efficiently *in vivo* than *in vitro*.

Recombination initiation by Spo11 is evolutionarily conserved⁴. We therefore tested for oligonucleotides bound to mouse SPO11 (mSPO11). An anti-mSPO11 antibody recognized 40.5 and 43.8 kDa polypeptides from adult testis (Fig. 2a, see arrowheads in lane 1), as expected for products of the major splicing isoforms *Spo11α* and *Spo11β* (refs 9, 10). Both were absent in the mock immunoprecipitation (Fig. 2a, lane 4) or with extracts from *Spo11*^{-/-} testes (Fig. 2a, lane 2). TdT-labelling of immunoprecipitates from wild-type testes yielded a smear of 47–59 kDa species (Fig. 2b, indicated by arrows in lanes 1 and 4). *Spo11*^{-/-} testes yielded only background, similar to a mock immunoprecipitation (Fig. 2b, compare lanes 2 and 5). The absence of labelled species from *Spo11*^{-/-} testes might result from spermatocyte death rather than lack of mSPO11 itself^{11,12}. To exclude this possibility, we also examined *Dmcl*^{-/-} testes, because DSBs are

formed and processed in the absence of DMC1, but strand exchange is impaired and spermatocyte apoptosis occurs at the same stage as in *Spo11*^{-/-} (refs 13, 14). As expected, labelled species were also isolated from *Dmcl*^{-/-} testes (Fig. 2b, lane 6). Mouse SPO11α was specifically reduced in *Dmcl*^{-/-} testes (Fig. 2a, lane 3), consistent with RNA expression data¹². Nevertheless, the mobility of mSPO11–oligonucleotide complexes was similar for *Dmcl*^{-/-} and wild-type testes samples (Fig. 2b, lanes 4 and 6). This result suggests that mSPO11α contributes little to DSB formation, consistent with the relatively late expression of *Spo11α* (ref. 12).

Oligonucleotide lengths from mice were analysed as above, except that the protein–DNA complexes were gel-purified *en masse* because they were not resolved as discrete bands. Two mSPO11-associated oligonucleotide populations (~12–26 nucleotides and ~28–34 nucleotides) were observed (Fig. 2c). No signal was obtained from a mock immunoprecipitation processed in parallel (Fig. 2c). The smaller oligonucleotide lengths were more heterogeneous compared to yeast, perhaps accounting for the lack of discrete bands on SDS–PAGE. Our findings reveal that the mechanism for processing Spo11-associated DSBs is evolutionarily conserved.

Spo11–DSBs are similar to topoisomerase II–DNA complexes formed after inhibition by compounds such as etoposide, in that both consist of phosphotyrosine linkages to 5' strand termini on both sides of a DSB^{15,16}. We hypothesized that the mechanism to process Spo11-associated DSBs also repairs topoisomerase II-induced DSBs. To test this idea, epitope-tagged topoisomerase II (Top2–HA3) was immunoprecipitated from vegetative yeast cultures and labelled with TdT. A species of appropriate size for a Top2–oligonucleotide complex was observed (Fig. 3a). This species was absent if cells expressed untagged Top2 (Fig. 3a). No difference in mobility of labelled and free Top2–HA could be discerned (Fig. 3b).

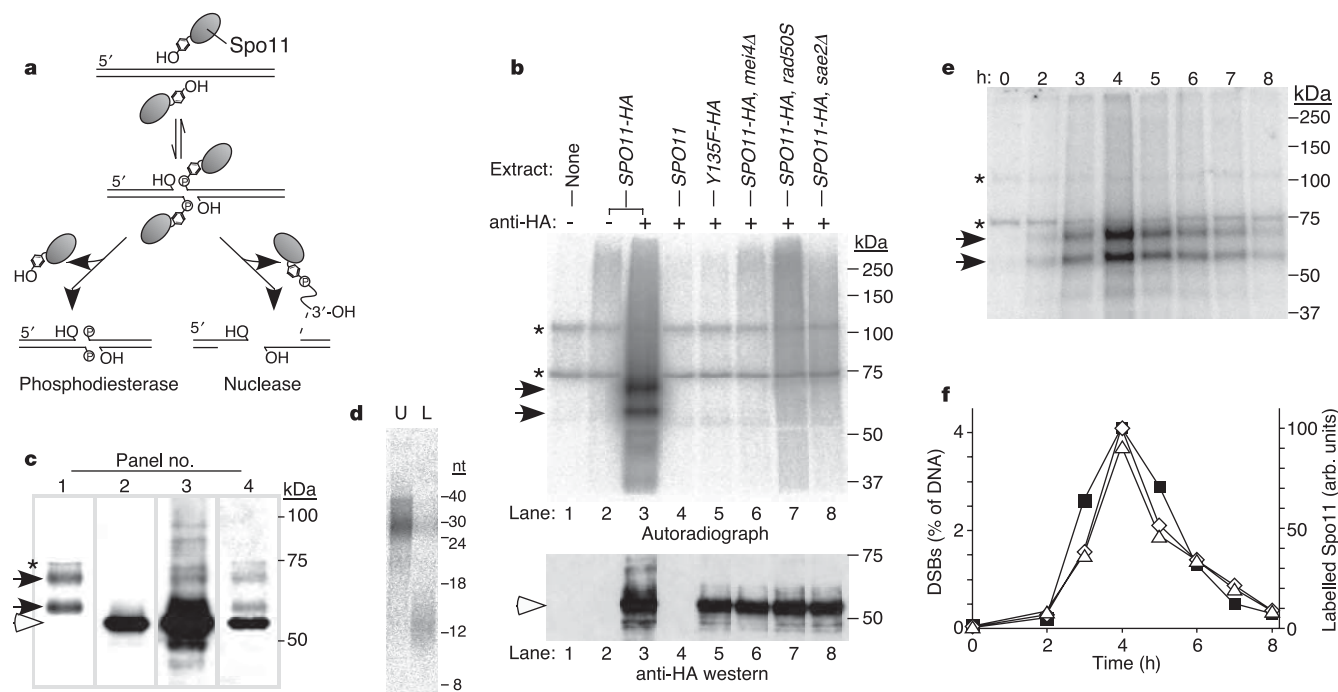


Figure 1 | Endonucleolytic processing of covalent Spo11–DSB complexes.

a, Alternative mechanisms for Spo11 release¹. **b**, Detection of Spo11–oligonucleotide complexes. Immunoprecipitates from the indicated genotypes with or without antibody were labelled with TdT. Shown are the autoradiograph (upper panel) and the anti-HA western blot (lower panel). **c**, Relative mobilities of free and oligonucleotide-associated Spo11–HA. Panels are as follows: autoradiograph (panel 1), low and high exposures of an anti-HA western blot (panels 2 and 3, respectively), re-exposure of the blot to film after partial fading of the chemiluminescent signal (panel 4). **d**, Sizes

of Spo11-associated oligonucleotides from the upper (U) and lower (L) SDS–PAGE bands. Size standards are indicated. **e**, Time course of appearance of Spo11–oligonucleotide complexes during meiosis. **f**, Quantification of upper (open diamond) and lower (open triangle) Spo11 species from **e**, and the DSB frequency at the *HIS4LEU2* recombination hotspot measured in the same culture (filled squares). Each point is a single measurement. Asterisks indicate non-specific labelling; closed arrows indicate Spo11-specific labelled species; open arrows indicate free Spo11–HA.

We estimated that we would have been able to detect a shift equivalent to 10 nucleotides or more. The absence of a detectable shift therefore suggests that Top2-associated oligonucleotides are smaller than for Spo11. These Top2 complexes are consistent with products of endonucleolytic processing of a low level of spontaneous topoisomerase-induced DSBs. However, it is also possible that these complexes arise from Top2-mediated cleavage adjacent to a pre-existing nick. Our experiments do not address whether Top2-associated DSBs might also be disjoined by tyrosine phosphodiester hydrolysis similar to the mechanism of Tdp1 phosphodiesterase¹⁷.

Mre11 was again a candidate for the nuclease because: (1) processing of topoisomerase lesions during bacteriophage T4 replication requires phage homologues of Rad50 and Mre11 (ref. 18); (2) the *Escherichia coli* homologue of the Rad50–Mre11 complex can remove protein complexes from duplex DNA ends¹⁹; and (3) mammalian MRE11 has been proposed to remove the protein covalently attached to the termini of the linear adenovirus genome²⁰. Surprisingly, however, the level of Top2–oligonucleotide complexes was not altered in *rad50S* or *sae2Δ* mutants (Fig. 3c), revealing that the formation of these complexes is not under the same genetic control as for Spo11. Therefore it is possible that Mre11 is not the relevant nuclease, these mutations affect the non-meiotic pathway differently or Mre11 is redundant with another nuclease(s). We also tested mutations eliminating known or suspected nucleases capable of cleaving 5′ single-stranded DNA at duplex–single-strand junctions (Rad2, Rad27, Yen1 and Din7)²¹. None affected the spontaneous

Top2–oligonucleotide complexes (Fig. 3c). Defining the genetic control for formation of these complexes remains an important challenge.

New details of the meiotic recombination pathway are suggested by characteristics of Spo11–oligonucleotide complexes, especially the presence of different oligonucleotide lengths. The mouse studies indicate that this heterogeneity is a conserved feature. It is possible that an initial longer form is converted by exonucleolytic activity to the shorter form, but a precursor–product relationship seems unlikely from the kinetics in yeast (Fig. 1f). More likely is that endonucleolytic cleavage can occur at either of two spacings from the DSB (Fig. 4). Several scenarios are possible. There might be two DSB classes that are processed symmetrically, with the two classes formed in equal numbers by chance. Alternatively, two cleavage spacings might occur randomly with equal probability, with spacing on one side independent of spacing on the other. In this model, long and short forms would be in equal amounts and there would be three DSB classes: symmetric long, symmetric short and asymmetric. Finally, the model we favour is that asymmetric cleavage at every DSB gives both a long and a short-oligonucleotide complex, providing a simple explanation for equal amounts of the two forms (Fig. 4).

Asymmetrically processed DSBs would have unexpected implications for the mechanism of recombination. Every recombination model posits that the two ends of a DSB behave differently during strand exchange: one end invades the homologous duplex and the

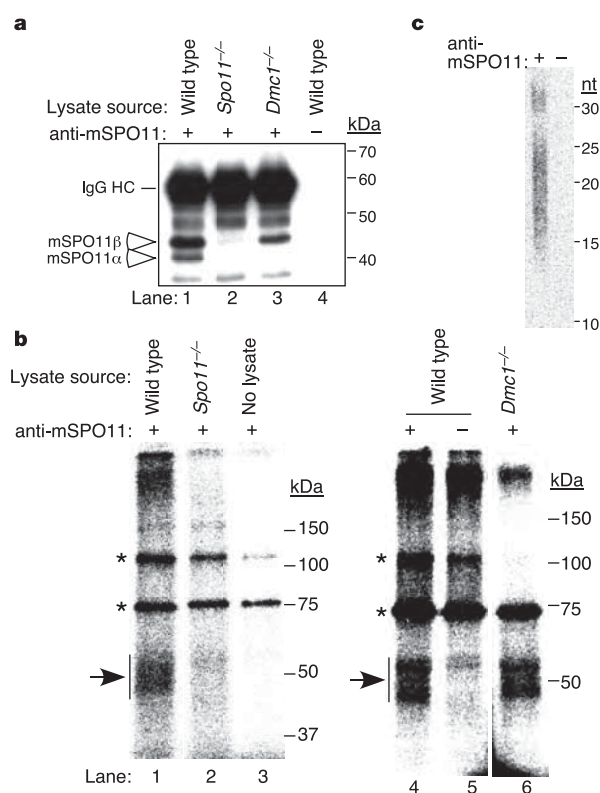


Figure 2 | SPO11-oligonucleotide complexes in mouse meiosis. **a**, Testis extracts immunoprecipitated with or without anti-mSPO11 were probed by western blotting with the same antibody. Each sample represents extract from an equivalent number of testes. IgG HC indicates immunoglobulin heavy chain. **b**, mSPO11-oligonucleotide complexes. mSPO11 was immunoprecipitated from testis extracts and labelled with TdT. Samples were from two testes (lanes 1–2) or six testes (lanes 4–6). Arrows indicate mSPO11-specific species; asterisks indicate non-specific bands. **c**, Sizes of mSPO11-associated oligonucleotides after gel purification and protease digestion. A mock immunoprecipitation lacking the anti-mSPO11 antibody was processed in parallel.

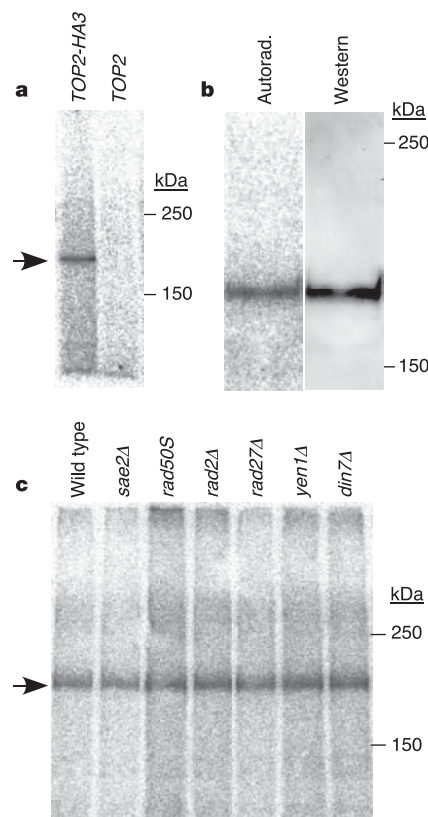


Figure 3 | Topoisomerase II-oligonucleotide complexes in non-meiotic yeast cells. **a**, Extracts were prepared from vegetatively growing yeast strains carrying topoisomerase II with (*TOP2-HA3*) or without (*TOP2*) an epitope tag. Samples were immunoprecipitated with anti-HA antibody and labelled with TdT. **b**, Comparison of electrophoretic mobilities of free and oligonucleotide-associated Top2-HA3. Immunoprecipitated Top2-HA3 was labelled with TdT, separated on 5% SDS-PAGE and transferred to PVDF membrane. The autoradiograph (Autorad.) and the anti-HA western blot (Western) of the membrane are shown. **c**, Formation of Top2–oligonucleotide complexes in nuclease-defective mutants. Arrow indicates the Top2-HA3-specific band.

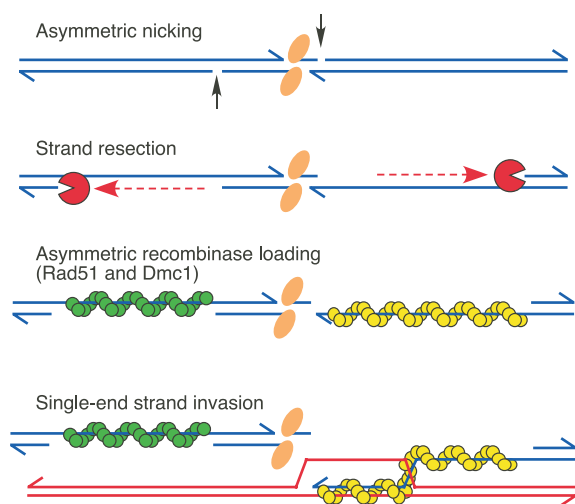


Figure 4 | Asymmetric steps in meiotic recombination. A Spo11 dimer (orange ellipses) creates a DSB which is processed by asymmetrically spaced nicks. Exonucleolytic resection initiates at these nicks, yielding single-stranded gaps. Rad51 and Dmc1 recombinases (green and yellow spiral chains, without specifying which) form nucleoprotein filaments on opposite sides of the DSB^{24,25}. Asymmetric strand invasion yields a stable strand exchange intermediate²⁴. Spo11–oligonucleotide complexes, stabilized by base pairing and protein–protein interactions, are proposed to remain associated with DSB ends until a step at or subsequent to strand exchange. See text for further details.

second is captured in a separate reaction^{22,23} (see Fig. 4). Analysis of intermediates in yeast supports this idea²⁴. Dmc1 and Rad51 have been proposed to bind to opposite ends of a DSB, with differences in these recombinases dictating asymmetric behaviour of the ends^{24,25} (Fig. 4). Our findings suggest that DSB ends may already be biochemically distinct at or before DSB processing, raising the possibility that subsequent asymmetric strand exchange is specified earlier than previously thought.

Temporal analysis of Spo11–DSB processing yields another unexpected insight. The kinetics for appearance and disappearance of both resected DSBs and Spo11–oligonucleotide complexes (Fig. 1f) were matched closely, implying that their lifespans are identical. Matched lifespans were unexpected because resected DSBs disappear when they give rise to strand exchange intermediates^{24,26}, whereas Spo11–oligonucleotide complexes presumably disappear on their degradation. It is possible that these distinct processes fortuitously occur with identical timing. However, matched lifespans could also be explained if turnover of Spo11–oligonucleotide complexes is mechanistically linked to turnover of resected DSBs. For example, Spo11–oligonucleotide complexes might remain associated with DSB ends until strand exchange, perhaps contributing further to asymmetry in the recombination reaction (Fig. 4).

METHODS

Yeast strains and culture methods. Meiotic cultures were prepared as described³. Vegetative cultures were grown in YPD medium. Strains for meiotic experiments are derivatives of SK1. Tagged *SPO11-HA3His6* (referred to for simplicity as *SPO11-HA*) and *spo11-Y135F-HA3His6* were described previously²⁷. Strains for topoisomerase experiments are derivatives of JN362a (ref. 28). Top2 protein was tagged at the carboxy-terminal with a triple HA repeat by a standard PCR fusion technique²⁹. Other mutations were introduced by one-step replacement and confirmed by Southern blot.

Yeast protein–oligonucleotide complexes. Cells were harvested from 100 ml of meiotic or 250 ml of vegetative culture, washed with water and lysed with glass beads in 15% ice-cold trichloroacetic acid (TCA). The TCA slurry was incubated on ice for 30 min before centrifugation (10 min, 16,000 g). Pellets were dissolved in 0.5 M Tris–Cl at pH 8.1, 2% SDS, 1 mM EDTA and 5% β-mercaptoethanol, boiled for 15 min and centrifuged at 16,000 g. Soluble protein was diluted tenfold

with IPD buffer (167 mM NaCl, 16.7 mM Tris–Cl at pH 8.1, 1.1 mM EDTA, 1.1% Triton X100 and 0.01% SDS) and incubated overnight at 4 °C with 2 μg monoclonal anti-HA antibody (clone F7; SantaCruz). Antibody complexes were captured with 20–50 μl protein–G–agarose beads (Roche) prewashed in IPD buffer. Immune complexes were collected by centrifugation, washed three times with IPD buffer, twice with × 1 NEB4 (50 mM K-acetate, 20 mM Tris-acetate, 10 mM Mg-acetate and 1 mM dithiothreitol at pH 7.9), resuspended and incubated 1 h at 37 °C in 50 μl × 1 NEB4 containing 0.5 mM CoCl₂, 50 units TdT (New England Biolabs) and either 10 μCi [α-³²P]-cordycepin triphosphate (5,000 Ci mmol⁻¹) for Spo11 or 50 μCi [α-³²P]-dCTP (6,000 Ci mmol⁻¹) for Top2. Immune complexes were then rinsed twice with IPD buffer and boiled 10 min in × 2 Laemmli buffer, then fractionated on SDS–PAGE. Gels were vacuum-dried and radiolabelled species were detected and quantified using film and Fuji phosphor screens and QuantityOne software (BioRad). For western analysis, protein was transferred to PVDF membrane in 10 mM CAPS at pH 11 and 10% methanol, probed with monoclonal anti-HA antibody conjugated to horseradish peroxidase (1:4,000 for Spo11–HA; 1:500,000 for Top2–HA; clone F7, SantaCruz,) and detected using ECL + reagent (Amersham).

To assess the size of Spo11-associated oligonucleotides, radiolabelled complexes were fractionated by 7.5% SDS–PAGE. Wet gels were aligned with the image from a phosphor screen, and the radiolabelled bands separately excised. Complexes were eluted by the crush and soak method, deproteinized for 2 h at 37 °C with 2 mg ml⁻¹ Pronase (Roche) in 100 mM Tris–Cl at pH 7.5, 0.5% SDS, 50 mM EDTA and 10 mM CaCl₂, then extracted twice with phenol/chloroform/isoamyl alcohol (25:24:1) and precipitated with 1 μg transfer RNA, ammonium acetate to 2.5 M and 2.5 volumes of ethanol. Precipitate was collected by centrifugation, vacuum dried, resuspended in 80% formamide, 10 mM EDTA and 0.1% xylene cyanol, boiled 2 min and chilled on ice. Samples were fractionated on an 18 cm-long sequencing gel (20% polyacrylamide [19:1], 7 M urea in × 1 TBE) at 16 mA for 70 min. Gels were fixed (10% methanol, 7% acetic acid and 5% glycerol) for 15 min and vacuum dried onto DE81 paper before autoradiography.

mSPO11–oligonucleotide complexes. *Spo11*^{-/-} and *Dmc1*^{-/-} mice were as previously described^{11,13}. Two to six testes from adult males (≥ 7 weeks old) were used for each experiment. Each testis was decapsulated and lysed in 500 μl 25 mM HEPES–NaOH at pH 7.9, 5 mM EDTA, 1% SDS, 2 mM DTT, 1 mM phenylmethylsulphonyl fluoride, leupeptin, pepstatin, chymostatin and aprotinin, each at 10 μg ml⁻¹. Lysates were heated at 95 °C for 10–20 min, chilled on ice and then centrifuged at 100,000 r.p.m. (355,040 g) for 15 min in a TLA100.2 rotor. Supernatants were diluted with 4 volumes of 1.25% Triton X-100, 187 mM NaCl, and 18.8 mM Tris–HCl at pH 8.0, then incubated with anti-mSPO11 antibody 129/180 (1 μg per testis; Kamiya Biomedical Company) at 4 °C overnight, followed by addition of 20–50 μl protein–A–agarose beads (Roche) and incubation for another 4 h. Beads were washed three times with IP buffer (1% Triton X-100, 150 mM NaCl, 15 mM Tris–HCl at pH 8.0 and 0.1% SDS) and twice with × 1 NEB4. Labelling was at 37 °C for 30 min using TdT and either [α-³²P] dCTP (Fig. 3b, lanes 4–6; 6,000 Ci mmol⁻¹) or [α-³²P]-cordycepin triphosphate (Fig. 3b, lanes 1–3; 5,000 Ci mmol⁻¹). Beads were washed twice with IP buffer, boiled in × 2 Laemmli sample buffer and fractionated on 8% SDS–PAGE.

To determine oligonucleotide lengths, mSPO11–DNA complexes from 14 testes were immunoprecipitated, labelled with [α-³²P] cordycepin triphosphate, separated on SDS–PAGE and transferred to PVDF membrane. The radiolabelled region containing mSPO11–DNA complexes and the corresponding region in the control, in which no antibody was added during the immunoprecipitation, were cut out. Oligonucleotides were eluted from the membrane by Pronase treatment, precipitated and separated on a sequencing gel as described above.

Received 24 February; accepted 3 June 2005.

- Keeney, S., Giroux, C. N. & Kleckner, N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375–384 (1997).
- Bergerat, A. et al. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386**, 414–417 (1997).
- Connelly, J. C. & Leach, D. R. Repair of DNA covalently linked to protein. *Mol. Cell* **13**, 307–316 (2004).
- Keeney, S. Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol.* **52**, 1–53 (2001).
- Alani, E., Padmore, R. & Kleckner, N. Analysis of wild-type and *rad50* mutants of yeast suggests an intimate relationship between meiotic chromosome synapsis and recombination. *Cell* **61**, 419–436 (1990).
- Diaz, R. L., Alcidi, A. D., Berger, J. M. & Keeney, S. Identification of residues in yeast Spo11p critical for meiotic DNA double-strand break formation. *Mol. Cell. Biol.* **22**, 1106–1115 (2002).

7. Liu, J., Wu, T.-C. & Lichten, M. The location and structure of double-strand DNA breaks induced during yeast meiosis: evidence for a covalently linked DNA-protein intermediate. *EMBO J.* **14**, 4599–4608 (1995).
8. Paull, T. T. & Gellert, M. The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. *Mol. Cell* **1**, 969–979 (1998).
9. Romanienko, P. J. & Camerini-Otero, R. D. Cloning, characterization, and localization of mouse and human SPO11. *Genomics* **61**, 156–169 (1999).
10. Keeney, S. *et al.* A mouse homolog of the *Saccharomyces cerevisiae* meiotic recombination DNA transesterase Spo11p. *Genomics* **61**, 170–182 (1999).
11. Baudat, F., Manova, K., Yuen, J. P., Jasin, M. & Keeney, S. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol. Cell* **6**, 989–998 (2000).
12. Romanienko, P. J. & Camerini-Otero, R. D. The mouse *Spo11* gene is required for meiotic chromosome synapsis. *Mol. Cell* **6**, 975–987 (2000).
13. Pittman, D. L. *et al.* Meiotic prophase arrest with failure of chromosome synapsis in mice deficient for Dmc1, a germline-specific RecA homolog. *Mol. Cell* **1**, 697–705 (1998).
14. Barchi, M. *et al.* Surveillance of different recombination defects in mouse spermatocytes yields distinct responses despite elimination at an identical developmental stage. *Mol. Cell. Biol.* (in the press).
15. Walker, J. V. & Nitiss, J. L. DNA topoisomerase II as a target for cancer chemotherapy. *Cancer Invest.* **20**, 570–589 (2002).
16. Fortune, J. M. & Osheroff, N. Topoisomerase II as a target for anticancer drugs: when enzymes stop being nice. *Prog. Nucleic Acid Res. Mol. Biol.* **64**, 221–253 (2000).
17. Pouliot, J. J., Yao, K. C., Robertson, C. A. & Nash, H. A. Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science* **286**, 552–555 (1999).
18. Stohr, B. A. & Kreuzer, K. N. Repair of topoisomerase-mediated DNA damage in bacteriophage T4. *Genetics* **158**, 19–28 (2001).
19. Connelly, J. C., de Leau, E. S. & Leach, D. R. Nucleolytic processing of a protein-bound DNA end by the *E. coli* SbcCD (MR) complex. *DNA Repair (Amst)* **2**, 795–807 (2003).
20. Stracker, T. H., Carson, C. T. & Weitzman, M. D. Adenovirus oncoproteins inactivate the Mre11–Rad50–NBS1 DNA repair complex. *Nature* **418**, 348–352 (2002).
21. Johnson, R. E., Kovvali, G. K., Prakash, L. & Prakash, S. Role of yeast Rth1 nuclease and its homologs in mutation avoidance, DNA repair, and DNA replication. *Curr. Genet.* **34**, 21–29 (1998).
22. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33**, 25–35 (1983).
23. Paques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).
24. Hunter, N. & Kleckner, N. The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell* **106**, 59–70 (2001).
25. Shinohara, M., Gasior, S. L., Bishop, D. K. & Shinohara, A. Tid1/Rdh54 promotes colocalization of Rad51 and Dmc1 during meiotic recombination. *Proc. Natl Acad. Sci. USA* **97**, 10814–10819 (2000).
26. Schwacha, A. & Kleckner, N. Identification of joint molecules that form frequently between homologs but rarely between sister chromatids during yeast meiosis. *Cell* **76**, 51–63 (1994).
27. Henderson, K. A. & Keeney, S. Tying synaptonemal complex initiation to the formation and programmed repair of DNA double-strand breaks. *Proc. Natl Acad. Sci. USA* **101**, 4519–4524 (2004).
28. Nitiss, J. L. *et al.* Amsacrine and etoposide hypersensitivity of yeast cells overexpressing DNA topoisomerase II. *Cancer Res.* **52**, 4467–4472 (1992).
29. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).

Acknowledgements We thank J. Nitiss for yeast strains; S. Schneider for assistance with strain construction; and M. Barchi, F. Cole, M. Di Giacomo and W. Mark for providing mice. Yeast work was supported by a grant from the National Institute of General Medical Sciences (to S.K.) and mouse work by a grant from the National Institute of Child Health and Human Development (to M. Jasin). M.J.N. is supported in part by a fellowship from the Human Frontiers Science Program.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.K. (keeney@mshcc.org).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

The lie of the land

Call it postdoc creep. In four years, the number of newly minted US geophysical PhDs who went on to become postdocs rose by nearly 50%, according to a survey conducted annually by the American Geophysical Union and the American Geological Institute. The survey, released this month, says that for the class of 1999, just under 40% opted to do a postdoc. Four years later, 58% of the class of 2003 decided to take a fellowship rather than a permanent position.

Meanwhile, as the number of geophysics postdocs increases, the gap between postdoc salaries and those for other positions remains wide. The median salary for a university postdoc — the lowest remunerated category — is about \$39,000 whereas for a newly minted PhD working in industry it is \$70,000.

The news for geophysicists isn't completely bad, though. The majority of 2003's graduates found work in their field, despite an economic downturn, with most reporting positive job satisfaction. And the percentage taking postdocs is still lower than it is for new PhDs in the life sciences, where a postdoc following a PhD has become the default.

But the growth in the number of geophysicists doing postdocs is a concern. More than scientists in other subdisciplines, geophysicists tend to work between receiving their undergraduate degree and starting their PhD, which means they tend to be more skilled and, theoretically, more likely to be hired without doing a postdoc.

So if fewer of them are finding permanent post-PhD positions, and there is still a disparity between postdoc and industrial salaries, there is less incentive for geophysicists with just an undergraduate education to go on for more training. The only way more permanent positions will open up — at least in government and academia — is with more funding from the US government for geophysics, an unlikely prospect, at best.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York
 345 Park Avenue South, 10th Floor,
 New York, NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
 Peter Bless
 Classified Sales Representatives
 Tel: +1 800 989 7718

**New York/Pennsylvania/
 Latin America:** Kelly Roman
**Midwest USA/Maryland/
 NIH:** Wade Tucker
East USA/Canada:
 Janine Taormina

San Francisco Office
Classified Sales Representative:
 Michaela Bjorkman
 225 Bush Street, Suite 1453
 San Francisco, CA 94104
 Tel: +1 415 781 3803
 Fax: +1 415 781 3805
 e-mail: m.bjorkman@naturesf.com

European Head Office, London
 The Macmillan Building,
 4 Crinan Street,
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director: Nevin Bayoumi (4978)
European Sales Manager: Andy Douglas (4975)

Advertising Production Manager: Billie Franklin
 To send materials use London address above.
 Tel: +44 (0) 20 7843 4814
 Fax: +44 (0) 20 7843 4966
 e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
 Patrick Phelan
 e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejp.com

Meeting (to move) up

Scientific conferences give new faculty members a chance to meet the leaders in their field and to give themselves some much-needed exposure. **Kendall Powell** works the room.

IMAGES.COM/CORBIS



It was a rookie mistake. In April, just months after establishing his own lab at the University of British Columbia in Vancouver, Leonard Foster strolled through the booths at the Keystone Symposium on Proteomics and Bioinformatics in Keystone, Colorado, chatting up friendly vendors during a break.

"I was slammed with meetings for two weeks after that, with people trying to get me to buy equipment," he recalls. He is wiser now to the different behaviour needed by a postdoc trawling for free T-shirts and by a new investigator hoping to make the right connections.

Attending scientific conferences early in the pre-tenure years can be critical for networking, recruitment and visibility. The informality and intimacy of smaller meetings usually affords the best opportunities for getting to know senior scientists in a field and for keeping up with major developments. There's no magic formula for becoming a prime schmoozer (nor is it even recommended), but people with experience can offer advice on making a good impression.

It's normal to feel overwhelmed and intimidated by the reputations of big names. Eleftherios Diamandis, a

cancer researcher at the University of Toronto, Canada, "felt like a lost chicken" at his first American Association for Cancer Research meeting some 15 years ago. But he struck up a conversation with someone whose badge carried a name he thought sounded Greek, like his own. That casual meeting resulted in one the longest and most fruitful collaborations of Diamandis's career.

Junior faculty members should not wait until they have a complete science story to go to meetings, says Dirk Schübeler, an epigenetics researcher at the Friedrich Miescher Institute in Basel, Switzerland. "You are cutting yourself off if you wait," he says. Not only do you miss a couple years of networking, but more importantly, he notes, you miss the chance to see how your research fits into the field.

Exposing yourself

The people you meet at conferences will be competitors, future collaborators, old friends and colleagues in the same boat, as well as grant and manuscript reviewers and writers of tenure letters. A newly established investigator should mingle with all of them, in time, to become known as a contributor of appealing work in the field. Working the crowd doesn't come naturally to most scientists, but senior faculty members recommend finding the right amount of boldness. That means stepping outside your comfort zone of mixing with friends and known colleagues.

"My agenda will fill itself up with people I already know, so you will definitely have to make the first move," says Ted Weinert, a cell-cycle expert at the University of Arizona in Tucson, who likes to talk science over a one-to-one lunch or coffee. "When people ask for my opinion on their research, I'm always flattered," he adds.

Bernard Golding, a veteran organic chemist at the University of Newcastle upon Tyne, UK, suggests approaching a senior scientist after his or her presentation with a follow-up question. "Most of us are happy to sit down with whomever. The more questions I get at a lecture, the more I like it," he says.

Others say it can ease introductions if they are made by a mutual friend. Or invite established colleagues to stop by your poster to view data that would interest them. Poster sessions offer an informal chance for a scientific chat.

Golding, who has chaired conferences on vitamin B12 chemistry, says organizers should try to be as inclusive of the younger generation as possible and not have a "clique of plenary speakers". At smaller conferences, he has included activities such as an eight-mile hike for people to get to know each other.

Also, he says, veteran scientists should make time to visit young scientists' posters — although the onus is on the presenters. "Put a massive amount of effort into the poster," Golding says. Don't cram too much on it, he warns, and make it eye-catching. An attractive poster can be the basis for the next presentation or a first publication, or can be hung outside an office to attract students (see *Nature* 434, 416–417; 2005). Colour handouts of your poster can serve as a permanent reminder of you and your work, he adds.

Don't spend the whole meeting chasing a few big names either. Leave time to mingle with other junior faculty members to commiserate about start-up problems, or find like-minded collaborators. For postdoc job candidates, this is a chance to get a foot in the door.



Question time: Bernard Golding says that quizzing big-shots about their lecture can provide a useful way to make contacts at a conference.

Morgan Tucker, a postdoctoral fellow at the University of Colorado, Boulder, spent much of his time at the international *Caenorhabditis elegans* meeting in Los Angeles in June probing new faculty members about their job searches. He says it was eye-opening to hear how friends' experiences at large and small universities compared.

As a postdoc, cell biologist Heike Fölsch sought out senior people at a Gordon Conference to discuss their science — critical for getting invited for job interviews. She started her own lab at Northwestern University in Evanston, Illinois in 2002. At July's Gordon Conference on molecular membrane biology in Andover, New Hampshire, she became an advocate for her lab and its publications "to present myself as somebody with interesting theories" in the field. She finds that small conferences give her more opportunities to make an impression than her field's annual large meeting, with thousands of participants.

Veterans agree that new faculty members should attend smaller series such as Gordon research conferences, Keystone symposia or summer meetings organized by the Federation of American Societies for Experimental Biology and those held at the University of Wales's Gregynog centre. Anything with more than 200 participants, and you get lost in the crowd, Weinert says.

Step right up

Smaller venues, on the other hand, have fewer slots for newcomers to give talks. One surefire way to obtain a booking is to submit stellar science, but there are ways to improve the odds.

Foster recommends asking the organizer for a talk. "It doesn't hurt and at least you've introduced yourself to one big name in the field," he points out. Organizers agree, but recommend doing it by e-mail, as a phone call puts them in an awkward spot. Another strategy is to tailor your abstract to the organizer's favourite topic.

"What rings my bell is to see someone doing something with a different spin on it," says Weinert. He says that the best abstracts for winning a talk slot



Scientific meetings offer an informal way to forge new collaborations and build a reputation.

emphasize uniqueness — be it a new approach, technique or theory — and broad appeal to the field.

Besides being great publicity for a lab's work, talks can help in recruiting. Even as a postdoc about to move, Schübeler began tacking a 'help wanted' job ad to the end of his talks. "Starting out, you are a blank page in terms of whether you are going to be able to run a lab," he says. "Your talk shows that you are in control and can manage a lab."

Getting the buzz

As a young investigator, Schübeler also hopes to get a scoop in his field. "You are still working out your projects, so it's good to have a feel of what's going on," he says. "It influences how you will publish if everyone thinks something is important to address." Fölsch notes that unpublished data get revealed at smaller, closed meetings, so these gatherings are best for keeping tabs on direct competitors.

Older scientists advise keeping low expectations and not to be too discouraged by feeling isolated at the first few meetings. Weinert suggests trying to meet only a couple of important contacts, and admits that he still feels intimidated at the start of a conference.

Now a field leader, Diamandis uses a large cancer meeting as an annual retreat for his entire lab group to celebrate a year of hard work. He notes that attending meetings consistently is the main way for researchers to move up among their peers, and to be invited to join committees and editorial boards.

Foster can see the wisdom in that. He may have made a rookie's mistake at the sales booths, but there's no mistaking the importance of attending meetings in his mind. "The biggest reason for meeting senior scientists," he says, "is that those are the people who are going to be reviewing your grants and your papers. If they know you, you'll have a better chance."

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

WEB LINKS

Keystone Symposia
www.keystonesymposia.org
 Gordon Research Conferences
www.grc.uri.edu
 Gregynog Centre
www.wales.ac.uk/gregynog
 FASEB summer research conferences
src.faseb.org

FIRST IMPRESSIONS



"Schmoozing? I'm not so good at that. But you can't get better advertising than giving a talk." — Leonard Foster, University of British Columbia, Vancouver, Canada

"I have the same feeling every time. It's too overwhelming, I'm intimidated, I get a little depressed, even. Then, in a couple of days, I'm in my element. It's an emotional rollercoaster." — Ted Weinert, University of Arizona, Tucson



"I only go to small meetings in the woods or mountains where you are bumping into each other all the time." — Dirk Schübeler, Friedrich Miescher Institute, Basel, Switzerland

"Go to smaller meetings in your field if you want to get tenured. The most influential people are not at the big meetings." — Heike Fölsch, Northwestern University, Evanston, Illinois

Corrections

In 'India's changing face' (*Nature* 436, 599; 2005), Jitu Mayor's name was misspelt as Major, and his informal first name was used. His formal first name is Satyajit. In 'Learning to mentor' (*Nature* 436, 435–436; 2005), a photo caption misidentifies Hopi Hoekstra as the woman pictured. In fact, the photo is of Anne Ephrussi, dean of graduate studies at the European Molecular Biology Laboratory.

MOVERS

David Wallace, director, Isaac Newton Institute for Mathematical Sciences, University of Cambridge



1994–current: Vice-chancellor, Loughborough University, Leicestershire, UK

1979–93: Tait Professor of Mathematical Physics, University of Edinburgh, UK

1972–79: Lecturer and reader, Department of Physics, University of Southampton, UK

1970–72: Harkness Fellow, Physics Department, Princeton University, New Jersey

David Wallace was nine or ten years old when his headmaster set the school class a simple problem, outside the usual curriculum. Wallace solved it, triggering a lifelong fascination with mathematics and science. His parents supported his experiments — even tolerating the hole burnt in his bedroom floor by some caustic soda he had secretly bought for his chemistry set.

Since then, he has dedicated his life to mathematics and science. That dedication recently culminated in his appointment as the next director of the Isaac Newton Institute for Mathematical Sciences in Cambridge. The institute supports programmes in theoretical, basic and applied mathematics.

Wallace is particularly looking forward to applying mathematics to environmental areas, for instance, by building models to examine factors that influence climate change. As well as the director's post at the institute, he will also serve as the N. M. Rothschild & Sons Professor of Mathematical Sciences.

After 12 years as an administrator, Wallace is very keen to return to the experimental world. "I always want to try something where I might fail," he says. "This is one of the things that make this position so challenging and rewarding."

His mentors, he says, always encouraged him to take on tough challenges. It's an admonition he took into his personal life as well as his career — he has completed four marathon runs. He understands the role of mentor very well: his PhD supervisor at the University of Edinburgh, he says, "taught us new techniques and inspired me in many ways".

He has also had to grapple with sometimes competing interests, serving on a number of corporate boards while also acting as university administrator — a situation he dealt with as if solving another mathematical problem. "Seeing how everything might fit together in a bigger picture can be very useful to resolve the apparently conflicting information one sometimes faces," he says.

He believes that instinct can sometimes be as important as logic when thinking of decisions and foresight. "You can always analyse data but in a lot of things you have to rely on instinct," he says.

The best advice he can give to young scientists is nevertheless very straightforward. "Do what you want to do," he says, "and take risks whenever you think it is right."

SCIENTISTS & SOCIETIES

Building a student network

In 2002, a dispute broke out between a group of students and Germany's Max Planck Society. It centred on disparities in the distribution of stipends and employment contracts for foreign and German students, and it had an unexpected spin-off. As a result of the argument, students at various Max Planck Institutes realized that they could use a more effective means of communicating with each other. This led to the formation of the Max Planck Phdnet, an online forum and networking site connecting more than 3,900 PhD students across the 78 institutes. Its goal is to foster communication on issues affecting students' professional lives.

Phdnet has evolved so that now PhD students from each institute elect a representative. These communicate with each other throughout the year on a web forum and also meet annually to discuss specific issues. The meetings are endorsed by the president of the Max Planck Society, who participates in some of them. This gives the students a chance to put questions directly to him. PhD representatives get an added incentive for volunteering, in that they learn about event management, networking and fund-raising.

Phdnet now promotes PhD advisory

committees, provides a comprehensive questionnaire to evaluate the quality of students' research experiences, and encourages students to make their voices heard in political decisions at the institutes. It has also begun to offer three types of seminars at various sites across Germany, exploring interdisciplinary subjects, career advice and 'soft' skills such as communication.

Because the Max Planck Institutes span life-science, chemical, physical, technological and humanities research, there is both the opportunity for broad interdisciplinary work and the challenge to communicate across these diverse areas. In the career seminars, Phdnet invites established researchers in various fields from academia and industry to talk about career paths. The soft-skills seminars stress effective scientific communication — specialists, give tips on effective oral and poster presentation and scientific writing.

Phdnet has grown quickly, but there are still a lot more things it can do. In a microbiologist's terms, it is only in its early log phase.

Benno Quade is spokesperson for Phdnet. Ajaybabu Pobbati is a team leader for the meetings work group.

▶ www.phdnet.mpg.de

GRADUATE JOURNAL

Master of my fate

In September, I will complete my graduate coursework and gain a Masters of Science degree. The next stage — DPhil, the Oxonian PhD — is now in sight.

But the vision of my scientific future is not becoming much clearer. A feeling of déjà vu dominated the past few months. It reminded me of the uncertainties I underwent when first considering what to do after medical training and, more recently, deciding to receive more academic training after earning an MSc.

Enrolling in research training in a lab and subject I don't know very well feels like a recurring theme in my life — making long-term commitments based on little information. So far, I think I've made good decisions, and remain on track, but I still harbour apprehension and doubts about the next move. Many factors come into play, such as thoughts about how promising the project is and if I will get along with the team. My analytical ego has no reservations about the latter point and the first one is unforeseeable. So I go for it. With slight discomfort.

A second issue should have cheered me up but instead helped me procrastinate; I received two offers. So, on top of my uncertainty about continuing with a PhD came the question of whom to work with. I felt bad about turning one bid down, especially because I liked the lab and the team as much as the one I accepted.

Now it's time to sit through this period of doubtfulness, and hope I will find out that I made the right decision again.

Tobias Langenhan is a first-year graduate student in neuroscience at the University of Oxford, UK.

The bell-curve drug

Choose your poison.

Jim Kling

Jonathan poked his head through my office door at the Institute.

"Feeling confident?" We were both Fellows and had known each other since we were Apprentices. I looked up from my computer screen showing the population-dynamics models, and smiled. We had both submitted proposals for the design of a sustainable park ecosystem with human hunters filling the role of top predator. We knew we were the front-runners. I might have stayed for some friendly banter, but I already had plans to see my father in the nursing home.

I took the sky train across town.

"How is your research coming on?" he asked me from the nursing-home bed, attended by a robotic nurse. I described the latest population models. I could see him struggling to comprehend.

"Can I help?" he asked, uncertainly. *No, you can't. Not any more.*

"It's under control," I said. I could see the relief in his eyes, then the pain.

"I wish...I could."

Aboard the sky train returning to the Institute, I met Rachel for the first time. She didn't have the vapid look of most of the other travellers, on their way to Carnival or some other entertainment. Most stared blankly at a TriVid programme showing on a screen in one corner of the car. I'd written the tragedy subroutine that had been used to create it.

She seemed full of purpose. I took her to be a Fellow at the Institute whom I'd somehow failed to meet before.

I was depressed, and she noticed. She asked why, and I told her about my father.

"It's the Neurypse, of course. He used to work at the Institute, too."

"Neurypse?"

Clearly, she wasn't a Fellow. "Johnson-MerckLilly developed it in the 2010s. It... makes us smarter," I said, deciding against a technical explanation.

"So you're better at your research. But it did that to your Dad?"

I nodded. "There are limits to how long

the mind can operate at peak efficiency. After that, it...declines."

"It sounds awful!"

I was taken aback. I swept my arm to indicate the pristine city visible outside the train window. Automated systems produced cheap and abundant energy. Farms produced, packaged and transported food, completely sustainably. Pollution was minimal. Few had to work — aside from those of us at the Institute, of course. "Neurypse made all this possible," I said.

"And your father?"



"He's 62. He's led a full life." I wondered if I believed it.

"It still sounds awful."

I was about to reply when the train slowed to a stop, and she shocked me by getting up. This was a district of whorehouses and opium dens. She handed me GPS coordinates on a piece of paper. "You should visit me," she said, and left the car amid three shuffling addicts.

The next day, I learned that Jonathan's proposal had won. He was gracious, even offering me a place on his implementation team. I reviewed my proposal and a sick feeling welled up in my stomach when I saw that I had omitted a crucial part of the experimental section. How could I have forgotten it? I thought of my father's blank stare, and felt a gnawing fear.

On the way to visit him a few days later, I saw Rachel again. She boarded at the same station she had exited, with a teenager in tow. His sunken eyes and pallor marked him as an addict. The boy saw the brief look pass between us. He started asking questions, and I told him about my work. His inquisitiveness surprised me. Rachel didn't look at me when I exited.

At the nursing home, my father had

taken a turn for the worse. For the first time, he did not recognize me. I left depressed. To clear my head, I decided to visit Rachel. The air cab dropped me off outside the city, in a village surrounded by animal pens and agricultural fields. There were no signs of mechanization. Rachel pushed through the crowd that had gathered to stare.

We walked to a bench near a copse of trees. Rachel told me about the village, how they grew their own food, made clothing, raised animals. She also told me

that the boy on the train was her son. He and his friends went to the opium dens frequently, but he stayed longer than most, wasting away for days until she came to retrieve him. "He can't find direction here. He doesn't see the importance of what we're doing," Rachel said.

"Importance?"

"We're reclaiming our humanity," she said. "The Institute's systems can clothe

and feed us, but they can't give us that."

I thought about the riders on the sky train, staring mindlessly at the TriVid show created by the subroutine I'd written. I looked around the village with a new interest. Rachel read my expression. "We're relearning everything. We still rely on the Institute to supplement us because we don't have it all down yet. You could help us." She paused briefly, gauging me. "Your father could stay here, too."

That decided it for me. He was wasting away in that grey-walled room, with nothing to do but watch banal TriVid programmes. Here at least he could walk a bit, and watch the goings-on. There would be no end of research to do, and I could stop taking the Neurypse because I wouldn't be fending off competition. A moment later, her son walked up to us, an eager expression on his face.

"May I ask you a question?" he said, looking at me.

"Of course."

"How do I join the Institute?"

Jim Kling writes about science and the future (and occasionally the past) from the shadows of a dormant volcano. More details can be found at <http://nasw.org/users/jkling>.

JACEY